A Section 508–conformant HTML version of this article is available at https://doi.org/10.1289/EHP12224.

Invited Perspective: Inroads to Biology-Informed Exposomics

Roel Vermeulen[1,2]

[1]Institute for Risk Assessment Sciences, Division of Environmental Epidemiology, Utrecht University, Utrecht, the Netherlands
[2]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands

https://doi.org/10.1289/EHP12224

Refers to https://doi.org/10.1289/EHP10239

The curse of dimensionality, a term often used to refer to interpretational difficulties that result from using high-dimensional data, is common to all omic fields. Exposomics—where new technologies (e.g., using omics, geospatial information, and sensors) are being developed rapidly to study exposures at scale—is no exception. The resulting large and rich data sets allow us to move from purely hypothesis-driven research toward more discovery-based studies where multiple exposures, either singly or in combination, are tested for association with a health phenotype. Unfortunately, such broad exploration of possible associations comes at a considerable price: multiplicity. The issue of multiplicity is not new, and over time several statistical approaches have been suggested that include either the use of more stringent *p*-value cutoffs or correcting the *p*-values for multiple comparisons.[1,2] Although these approaches provide opportunities for controlling false-positive error rates, they, too, come at a cost: reduced sensitivity and lower statistical power to detect true positives. So is there a better way forward?

My colleagues and I previously wrote about using biological information in mapping the exposome to health.[3] We argued that one could take a pragmatic view as to which exposures to prioritize in an exposome study, namely: *a*) the ones we are exposed to and are therefore detectable in the body, and *b*) the ones that we know can induce biological effects. Including biological knowledge in our exposome-health analyses can be achieved in multiple ways. One technological solution is to perform biology-based measurements by characterizing the molecular interaction between an exposure and a functional biomolecule. After adding a biofluid to a microwell plate coated with a receptor with a known function, any captured biologically active ligand can be extracted and characterized chemically.[3] Another avenue is to include relevant prior biological knowledge and associated probabilities ("priors") in the statistical analyses.

The paper by Nakiwala et al. in the current issue of *Environmental Health Perspectives* is an example of this second approach.[4] They combined information on adverse outcome pathways (AOPs) with results from *in vitro* high-throughput screening assays in the ToxCast/Tox21 database to preselect detectable exposures thought to be relevant to thyroid hormone homeostasis. This resulted in reducing the number of exposures under consideration from 16 to 13 bioactive chemicals. Although the drop in numbers turned out to be marginal for their specific application, this is likely at least partly due to the more limited set of chemicals (from only two chemical classes) that they considered. When using broad chemical screens across many chemical classes, the reduction is likely to be much more substantial.

Nakiwala et al. used the biological knowledge as a prior and argued there was no need for multiple testing correction of the *p*-values they chose for calling out significance. Including biological priors in exposome (wide) analyses makes a concrete inroad into biology-informed (i.e., functional) exposomics and may help to improve the trade-off between sensitivity and false-positive rates. However, even though Nakiwala et al. did not describe their prior in large detail and included it only indirectly in their analyses, this prior should still be subject to the same critical evaluation and discussion as one would do with priors used in a more formal (i.e., Bayesian) analysis. To better understand the utility of the approach proposed by Nakiwala et al., we thus need to ask two questions: How informative (strong) is this prior? And does it seem to be compatible with the observed data, that is, is it appropriate?

So how strong is the prior? By arguing that no multiple testing correction was needed given prior biological information, Nakiwala et al. implicitly awarded exposures for which they have this information a 16- to 96-fold (16 exposures and 6 end points) higher probability of selection over that of exposures for which this information was not available or that tested negative—assuming they would have otherwise used a Bonferroni correction and depending on whether the scope of the correction was outcome-wide or not. This seems a rather strong prior and even more so because it was also used to completely exclude chemicals from further analyses.

Is this prior appropriate? Assuming the study was adequately powered, one would expect that any false null hypothesis would be rejected with a probability of $\sim 80\%$. However, of the 78 *p*-values reported in their Table 6 for dose–response relationships between all 13 selected exposures and 6 thyroid outcomes, only 3 passed the conventional threshold of $p < 0.05$ for calling significance. This number is well within the range that one would expect under the complete null, which would suggest that either the prior was overly optimistic or the study was underpowered even after using the prior information.

Although the paper by Nakiwala et al. does not provide formal statistical proof that the prior they used, based on *in vitro* assays and AOPs, informed the presented epidemiological analyses (i.e., more precise estimates of the dose–response or increased certainty on the presence or absence of associations), it is an important paper. It sparks a discussion on how to *a*) construct informative priors and what evidence bases can be used for this, *b*) integrate priors more formally in the statistical analyses, and *c*) properly evaluate the results of biologically informed models. We now tend to rely too often on post hoc reasoning, where we search for a biological explanation after having obtained the results of statistical tests. This process is prone to overinterpretation. However, using a prior without evaluating its impact on the analysis or whether it is appropriate runs the risk of the same overinterpretation.

The roadmap for how to best integrate biological information in exposome studies has not been written yet. It is evident, however,

Address correspondence to Roel Vermeulen, Institute for Risk Assessment Sciences, Division of Environmental Epidemiology, Utrecht University, Yalelaan 2, 3584 CM Utrecht, the Netherlands. Email: r.c.h.vermeulen@uu.nl

The author declares he has nothing to disclose.

Received 2 October 2022; Revised 7 October 2022; Accepted 7 October 2022; Published 9 November 2022.

that there is an urgent need for such a roadmap so as to improve inference and reproducibility of exposome studies. In this we need to have a pluralistic perspective exploring technical, experimental, and statistical solutions alike. As in many other areas in life, the road is likely to be slow, with a fair number of twists and turns. But we must start this journey if we are to learn important lessons.

Acknowledgments

The author gratefully acknowledges the help and critical discussions with L. Portengen (Utrecht University).

R.V. is funded by EXPOSOME-NL, the exposome research consortium of the Netherlands [Dutch Research Council (NWO); project no. 024.004.017] and EXPANSE (EU-H2020; grant no. 874627).

References

1. Streiner DL, Norman GR. 2011. Correction for multiple testing: is there a resolution? Chest 140(1):16–18, PMID: 21729890, https://doi.org/10.1378/chest.11-0523.
2. Greenland S. 2017. Invited commentary: the need for cognitive science in methodology. Am J Epidemiol 186(6):639–645, PMID: 28938712, https://doi.org/10.1093/aje/kwx259.
3. Chung MK, Rappaport SM, Wheelock CE, Nguyen VK, van der Meer TP, Miller GW, et al. 2021. Utilizing a biology-driven approach to map the exposome in health and disease: an essential investment to drive the next generation of environmental discovery. Environ Health Perspect 129(8):85001, PMID: 34435882, https://doi.org/10.1289/EHP8327.
4. Nakiwala D, Noyes PD, Faure P, Chovelon B, Corne C, Gauchez AS, et al. 2022. Phenol and phthalate effects on thyroid hormone levels during pregnancy: relying on *in vitro* assays and adverse outcome pathways to inform an epidemiological analyses. Environ Health Perspect 130(11):117004, https://doi.org/10.1289/EHP10239.