Environ Health Perspect

DOI: 10.1289/EHP4817

Supplemental Material

Privacy Risks of Sharing Data from Environmental Health Studies

Katherine E. Boronow, Laura J. Perovich, Latanya Sweeney, Ji Su Yoo, Ruthann A. Rudel, Phil Brown, and Julia Green Brody

Table of Contents

Table S1. Members of the Advisory Council that convened in 2013 and evaluated the selection of environmental health studies highlighted in this article.

Name	Current Affiliation and Related Experience
Phil Bereano, PhD	Professor Emeritus, Technology and Public Policy, University of Washington, Seattle, WA, USA Bereano chaired the ACLU Committee on Data Collection, Storage, and Dissemination and has worked with many US and international agencies on science and technology policy.
Elaine Cohen Hubal, PhD	Senior Science Advisor, US Environmental Protection Agency, Research Triangle Park, NC, USA Cohen Hubal directed ExpoCast in the National Center for Computational Toxicology, generating publicly-available exposure data and computational tools for chemical prioritization and risk assessment.
Dean Gallant	Senior Consultant, HRP Consulting Group, Lake Success, NY, USA Gallant was formerly the Assistant Dean for Research Policy at Harvard University and is a specialist in data privacy.
Bob Gellman, JD	Privacy and Information Policy Consultant, Washington, D.C., USA Gellman has worked extensively on federal privacy policy. He chaired the Privacy and Confidentiality subcommittee of the National Committee on Vital and Health Statistics and worked on the Freedom of Information Act.
Ian Kerr, LLB, PhD	Full Professor and Canada Research Chair in Ethics, Law, and Technology, University of Ottawa, Ottowa, Ontario, Canada Kerr holds appointments in law, medicine and philosophy and information studies at the University of Ottawa. His work, *Lessons from the Identity Trail*, focuses on how information and authentication technologies affect identity and right to be anonymous, and was supported by the Social Sciences and Humanities Research Council.
Nancy M.P. King, JD	Professor, Social Sciences and Health Policy and Institute for Regenerative Medicine, Wake Forest School of Medicine, Winston Salem, NC, USA King has written extensively on informed consent and currently focuses on genetics and biobanking. She is a Fellow of the Hastings Center and directs the Wake Forest University Research Ethics Consultation Program.

Name	Current Affiliation and Related Experience
Sheldon Krimsky, PhD	Professor, Urban and Environmental Policy and Planning, Tufts University, Medford, MA, USA Krimsky is an Elected Fellow of the AAAS, a leader in bioethics, and founder of the Council for Responsible Genetics, which champions genetic privacy.
Karen Miller	President, Huntington Breast Cancer Action Coalition, Huntington, NY, USA Miller is founder of Huntington (Long Island) Breast Cancer Action Coalition and a long-time breast cancer activist who has served in advisory roles for NIEHS and NCI.
Arvind Narayanan, PhD	Associate Professor, Computer Science, Princeton University, Princeton, NJ, USA Narayanan has extensive technical experience in de-anonymizing, including prominent work on social network datasets, large sparse datasets, and fingerprinting.
David Ozonoff, MD, MPH	Professor and Chair Emeritus, Environmental Health, Boston University School of Public Health, Boston, MA, USA Ozonoff is an experienced environmental health researcher with interests in mathematics and smoothing of geographic data to protect privacy.
Laura Perovich	Ph.D. candidate, MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA Perovich's experience analyzing Silent Spring Institute Household Exposure Study data contributed to the development of this study.
Dale Sandler, PhD	Senior Investigator, Epidemiology Branch and Chronic Disease Epidemiology Group, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA Sandler is Chief of the NIEHS Epidemiology Branch and PI of large EH cohort studies, including the Sister Study, Agricultural Health Study, and GuLF Study, leading to her interest in data-sharing and privacy.
Jessica Tovar	Organizer, Local Clean Energy Alliance, Oakland, CA, USA Tovar collaborated with Silent Spring Institute as a community organizer at Communities for a Better Environment, an environmental justice organization. She is trained in human subjects ethics and conducted home visits to collect data for the HES.

Table S2. Chemicals ranked by highest absolute loadings on the first principal component resulting from principal component analysis of air and dust measurements in the Household Exposure Study.

Sample matrix and rank	Original reporting limits		Censored reporting limits	
	Chemical	Loading	Chemical	Loading
Air				
1	Heptachlor	0.42	o-Phenylphenol	0.65
2	o-Phenylphenol	0.39	Heptachlor	0.34
3	gamma-Chlordane	0.39	gamma-Chlordane	0.31
4	alpha-Chlordane	0.36	Propoxur	0.28
5	Diazinon	0.35	alpha-Chlordane	0.27
6	Chlorpyrifos	0.28	Chlorpyrifos	0.25
7	PCB 52	0.24	Benzyl butyl phthalate	0.22
8	Propoxur	0.24	Diethyl phthalate	0.17
9	Benzyl butyl phthalate	0.20	Di-n-butyl phthalate	0.16
10	Bis(2-ethylhexyl) adipate	-0.13	Bis(2-ethylhexyl) adipate	-0.16
11	Diethyl phthalate	0.09	Diazinon	0.12
12	Di-n-butyl phthalate	0.09	PCB 52	0.07
13	Diisobutyl phthalate	-0.07	Diisobutyl phthalate	-0.06
Dust				
1	Chlorpyrifos	0.30	Piperonyl butoxide	0.45
2	Diazinon[a]	0.28	Chlorpyrifos	0.33
3	PCB 105[a]	0.27	trans-Permethrin	0.33
4	Benz(a)anthracene	0.27	cis-Permethrin	0.30
5	PCB 153[a]	0.26	Benz(a)anthracene	0.25
6	PCB 52[a]	0.26	Carbaryl	0.25
7	gamma-Chlordane	0.26	Benzo(a)pyrene	0.23
8	Benzo(a)pyrene	0.25	Propoxur	0.22
9	alpha-Chlordane	0.25	Methoxychlor	0.21
10	Methoxychlor	0.24	gamma-Chlordane	0.20
11	Piperonyl butoxide	0.23	alpha-Chlordane	0.2
12	4,4'-DDD[a]	0.22	Diethyl phthalate	0.17
13	Chlorothalonil[a]	0.22	Bis(2-ethylhexyl) adipate	0.17
14	Carbaryl	0.21	Benzyl butyl phthalate	0.16
15	4,4'-DDE[a]	0.2	Di-n-butyl phthalate	0.15
16	Propoxur	0.14	4,4'-DDT	0.14
17	Benzyl butyl phthalate	0.11	Di-n-hexyl phthalate	0.12
18	4,4'-DDT	0.11	Diisobutyl phthalate	0.02
19	Diethyl phthalate	0.07		
20	trans-Permethrin	0.07		
21	Di-n-hexyl phthalate	0.07		
22	Di-n-butyl phthalate	0.06		

23	cis-Permethrin	0.06
24	Diisobutyl phthalate	-0.05
25	Bis(2-ethylhexyl) adipate	0.05

Note: Principal component analysis was used to help identify chemicals that were influential in the K-means clustering analysis of the same data. For the analyses of air data with original reporting limits (Figure 1B), air data with censored reporting limits (Figure 1B), and dust data with original reporting limits (Figure 1C), the K-means clusters correspond to regional subgroups that diverge along PC1, with Massachusetts homes having higher PC1 scores than California homes. The analysis of dust data with censored reporting limits (Figure 1D) did not produce accurate K-means clusters by region, with homes from Massachusetts and California overlapping along PC1 and PC2.

[a]Chemical was not included in the dust analysis with censored reporting limits because it no longer met the minimum detection frequency.

Table S3. Chemicals ranked by highest absolute loadings on the first principal component resulting from principal component analysis of air measurements in the Green Housing Study.

Sample matrix and rank	Original reporting limits		Constant reporting limits	
	Chemical	Loading	Chemical	Loading
Air				
1	Musk ketone	0.37	Musk ketone	0.38
2	Tris(1,3-dichloroisopropyl) phosphate	0.37	Musk xylene	0.37
3	Musk xylene	0.36	Tris(1,3-dichloroisopropyl) phosphate	0.36
4	BDE 47	0.30	BDE 47	0.29
5	BDE 100	0.26	BDE 100	0.26
6	BDE 99	0.25	BDE 99	0.25
7	BDE 28	0.23	BDE 28	0.23
8	Tonalide	0.21	Tonalide	0.21
9	Methyl paraben	0.20	Methyl paraben	0.20
10	Galaxolide	0.20	Galaxolide	0.19
11	Diisononyl phthalate	0.17	Diisononyl phthalate	0.17
12	1,3-dichloro-2-propanol	0.16	1,3-dichloro-2-propanol	0.16
13	Triphenyl phosphate	0.14	Triphenyl phosphate	0.14
14	Butylbenzyl phthalate	0.14	Butylbenzyl phthalate	0.13
15	Bis(2-ethylhexyl) phthalate	0.14	Bis(2-ethylhexyl) phthalate	0.13
16	Diethyl phthalate	0.13	Diethyl phthalate	0.13
17	Benzophenone-3	0.12	Benzophenone-3	0.12
18	Triclosan	0.11	Tris(2-chloroethyl) phosphate	0.11
19	Bis(2-ethylhexyl) adipate	0.11	Triclosan	0.11
20	Tris(2-chloroethyl) phosphate	0.11	Bis(2-ethylhexyl) adipate	0.11
21	Benzophenone	0.09	Benzophenone	0.09
22	4-t-nonylphenol	0.07	4-t-nonylphenol	0.07
23	Tris(1-chloro-2-propyl) phosphate	0.04	Tris(2-butoxyethyl) phosphate	0.04
24	Tris(2-butoxyethyl) phosphate	0.03	2-ethylhexyl 2,3,4,5-tetrabromobenzoate	0.04
25	2-ethylhexyl 2,3,4,5-tetrabromobenzoate	0.02	Tris(1-chloro-2-propyl) phosphate	0.04
26	Dicylcohexyl phthalate	-0.02	2,3-dibromo-1-propanol	0.01
27	Di-n-butyl phthalate	-0.01	Dicylcohexyl phthalate	-0.01
28	2,3-dibromo-1-propanol	0	Di-n-butyl phthalate	-0.01

Note: Principal component analysis was used to help identify chemicals that were influential in the K-means clustering analysis of the same data. For the analyses of air data with original reporting limits (Figure 1E) and censored reporting limits (Figure 1F), the K-means clusters correspond to regional subgroups that diverge along PC1, with Cincinnati homes having higher PC1 scores than Boston homes.