Taylor & Francis
Taylor & Francis Group

Accountability in Research
Policies and Quality Assurance

ISSN: 0898-9621 (Print) 1545-5815 (Online) Journal homepage: http://www.tandfonline.com/loi/gacr20

Pesticides, Neurodevelopmental Disagreement, and Bradford Hill's Guidelines

Kristin Shrader-Frechette Ph.D. & Christopher ChoGlueck B.S., B.A.

To cite this article: Kristin Shrader-Frechette Ph.D. & Christopher ChoGlueck B.S., B.A. (2017) Pesticides, Neurodevelopmental Disagreement, and Bradford Hill's Guidelines, Accountability in Research, 24:1, 30-42, DOI: 10.1080/08989621.2016.1203786

To link to this article: http://dx.doi.org/10.1080/08989621.2016.1203786

Accepted author version posted online: 27 Jun 2016.
Published online: 27 Jun 2016.

ACCOUNTABILITY IN RESEARCH
2017, VOL. 24, NO. 1, 30–42
http://dx.doi.org/10.1080/08989621.2016.1203786

Taylor & Francis
Taylor & Francis Group

Pesticides, Neurodevelopmental Disagreement, and Bradford Hill's Guidelines

Kristin Shrader-Frechette, Ph.D.[a] and Christopher ChoGlueck B.S., B.A.[b]

[a]Departments of Biological Sciences and Philosophy, University of Notre Dame, Notre Dame, Indiana, USA; [b]Department of History and Philosophy of Science & Medicine, Indiana University, Bloomington, Indiana, USA

ABSTRACT

Neurodevelopmental disorders such as autism affect one-eighth of all U.S. newborns. Yet scientists, accessing the same data and using Bradford-Hill guidelines, draw different conclusions about the causes of these disorders. They disagree about the pesticide-harm hypothesis, that typical United States prenatal pesticide exposure can cause neurodevelopmental damage. This article aims to discover whether apparent scientific disagreement about this hypothesis might be partly attributable to questionable interpretations of the Bradford-Hill causal guidelines. Key scientists, who claim to employ Bradford-Hill causal guidelines, yet fail to accept the pesticide-harm hypothesis, fall into errors of trimming the guidelines, requiring statistically-significant data, and ignoring semi-experimental evidence. However, the main scientists who accept the hypothesis appear to commit none of these errors. Although settling disagreement over the pesticide-harm hypothesis requires extensive analysis, this article suggests that at least some conflicts may arise because of questionable interpretations of the guidelines.

KEYWORDS

Bradford-Hill; cause; neurodevelopment; pesticide; statistical significance; trimming the data

Introduction

Roughly 500,000 of the 4 million children born annually in the United States have neurodevelopmental disorders such as autism or attention-deficit-hyperactivity disorder (Landrigan, Lambertini, and Birnbaum, 2012). Nevertheless some scientists, who have access to the same evidence and use the same causal guidelines to evaluate it, draw opposed conclusions about the pesticide-harm hypothesis—that typical U.S. prenatal pesticide exposure can cause some neurodevelopmental problems. Given the potential to prevent some of these problems, scientific conflict over the hypothesis warrants assessment.

CONTACT Kristin Shrader-Frechette kshrader@nd.edu Departments of Biological Sciences and Philosophy, University of Notre Dame, 100 Malloy Hall, Notre Dame, IN 46556, USA.

Scientific disagreement about pesticide effects

In 2011, three different research teams from Berkeley, Columbia, and Mt. Sinai medical schools independently published results supporting the hypothesis of pesticide-induced neurodevelopmental harm (Bouchard et al., 2011; Engel et al., 2011; Rauh et al., 2011; see Grandjean and Landrigan, 2014). Together they won the 2012 "Paper of the Year" award from *Environmental Health Perspectives* (EHP). However, during 2011–2013 three other groups of scientists, claiming to use Bradford-Hill causal guidelines, failed to accept the pesticide-harm hypothesis; they claimed "evidence of causality … is not compelling" (Burns et al., 2013, pp. 127, 261; Li et al., 2012; Mink, Kimmel, and Li, 2012), that "epidemiologic studies do not support a causal association" (Li et al. 2012, p. 174; see Mink, Kimmel, and Li, 2012, p. 312). Which group of scientists is right?

The EHP-award-winning scientists who accept the pesticide-harm hypothesis showed that different studies/methods confirm that as children's *in-utero* exposure to the lowest-dose-organophosphate pesticides increases, so do later deficits in areas such as IQ, working memory, and perceptual reasoning (e.g., Bouchard et al., 2011; Engel et al., 2011; Rauh et al., 2011). Each 10-fold increase in a pregnant woman's organophosphate levels is associated with her child's IQ dropping 6 points by age 7 (Bouchard et al., 2011). For each standard-deviation-increased exposure to the organophosphate chlorpyrifos, measured in umbilical-cord-blood plasma, childhood IQ declines 1.4%, and working memory 2.8%, both by age 7 (Rauh et al., 2011). The greater the prenatal-organophosphate exposure, the poorer the cognitive development; effects begin at 12 months and continue through childhood (Engel et al., 2011).

However scientists who are critical of the pesticide-harm hypothesis say many specific pesticide-neurodevelopmental associations have been neither experimentally tested nor replicated (Burns et al., 2013, pp. 255–256); are not "statistically significant"; and reveal "no consistent patterns of adverse association" (Li et al., 2012, p. 132; see Mink, Kimmel, and Li, 2012, p. 312). Therefore, they claim poverty or air pollution, not pesticides, likely explains observed neurodevelopmental harm (Burns et al., 2013, esp. p. 261; see Li et al., 2012, p. 174; Mink, Kimmel, and Li 2012, p. 312).

Although full analysis of the controversy over the pesticide-harm hypothesis would require meta-analysis and full evaluation of all data and issues involved, this article addresses only one aspect of this controversy, that is, how well the respective scientists employ the Bradford Hill causal guidelines that they claim to use to defend their positions. Subsequent paragraphs suggest that part of the reason these researchers disagree may be the questionable Burns-Li-Mink interpretations of Bradford-Hill causal guidelines (BHG).

Discussion

Bradford Hill's guidelines

Austin Bradford Hill's (1965) nine "viewpoints" or guidelines for assessing causal associations—strength, consistency, specificity, temporality, biological gradient, plausibility, coherence, experimental manipulation, and analogy—are foundational in many statistics-related disciplines (e.g., Cranor, 2006). Indeed, the U.S. Environmental Protection Agency (U.S. EPA, 2005, 2010) mandates use of BHG in causal assessment, especially in "weight-of-evidence " or "preponderance-of-the-information" strategies requiring considering "all available data" (U.S. EPA, 2010, p. 27, 2011, p. 27). As outlined by Hill (1965) and the U.S. EPA (2005):

(1) The *strength* of association BHG refers to the "excess" of some effect, regularly associated with a possible cause, such as the ten-fold increase in lung cancer for tobacco smokers, compared with nonsmokers (Hill, 1965, pp. 295–296). However, *strength* requires taking account of confounders (U.S. EPA, 2005, pp. 2–12; Hill, 1965, p. 296), as a "modest[-sized] risk … does not preclude causality" but may reflect conditions such as lower-exposure levels (U.S. EPA, 2005, pp. 2–13; see Hill, 1965, p. 296).
(2) The *consistency* BHG refers to the fact that a putative "observed association" between cause and effect has been "repeatedly observed by different persons, in different places, circumstances, and times," and by "independent studies" (Hill, 1965, pp. 296–297). Nevertheless "discordant effects" are not evidence of inconsistency if strong reasons explain the discordance (U.S. EPA 2005, p. 2–13).
(3) The *specificity* BHG refers to the fact that often an association ties a putative cause to "specific" populations of victims who have certain "types of disease," such as chimney sweeps with excess scrotal cancer (Hill, 1965, p. 297). However, specificity is not required for causality, as diverse agents often cause single diseases, and one agent often causes different effects (U.S. EPA 2005, pp. 2–13, 2–14).
(4) The *temporality* BHG refers to the fact that a cause must precede an effect in time (Hill, 1965, pp. 297–298), though "harm may be latent for "20 years or longer" (U.S. EPA, 2005, pp. 2–14).
(5) The *biological gradient* BHG refers to the fact that stronger/larger causes produce stronger/larger effects, as when heavier smokers have higher lung-cancer rates (Hill, 1965, p. 298).
(6) The *plausibility* BHG refers to the fact that a putative cause-effect association relies on current "biological knowledge," such as established theory or underlying mechanisms—though evidence often is unavailable in new scientific areas (Hill, 1965, p. 298). The U.S. EPA (2005, pp. 2–12, 2–14) recommends assessing possible cause/effect

modes-of-action by evaluating "plausible biological mechanisms to explain" them.(7) The *coherence* BHG refers to the facts that causal associations must "not seriously conflict with the … biology of the disease" (Hill, 1965, p. 298), and are stronger when "other lines of evidence," e.g., animal data, support them (U.S. EPA, 2005, p. 2–14).

(7) The *experimental manipulation* BHG refers to the fact that a causal hypothesis can gain/lose support if certain conditions, such as "preventive action," change the "frequency of the associated events," while other factors remain constant (Hill, 1965, p. 298). Such manipulations—"experimental or semi-experimental evidence" of causality—are the "strongest support for the causation hypothesis" (Hill, 1965, pp. 298–299; see U.S. EPA, 2005, pp. 2–14).

(8) The *analogy* BHG refers to the fact that purported causal associations gain strength from "similar evidence" of similar effects (Hill, 1965, p. 299), e.g., similar "modes of action" or "structural [or chemical] analogues" of different causal agents (U.S. EPA, 2005, pp. 2–15).

Three prominent errors

Although causal assessments can go wrong in various ways, Hill, the U.S. EPA, and others warn of three prominent failures in using BHG or weight-of-evidence guidelines—all of which appear to play a role in the pesticide-harm-hypothesis controversy. These are errors of trimming the guidelines, requiring statistically-significant data, and ignoring semi-experimental evidence.

The guideline-trimming *error involves begging causal questions by ignoring one or more of the BHG*, such as the plausibility (BHG 6) of a causal connection, given potential mechanisms or modes-of-action. To avoid guideline-trimming, both the U.S. EPA (2005, pp. 2–11ff) and Hill (1965, p. 299) warn there are nine guidelines, "from all of which we should study association before we cry causation"; no single consideration "can bring indisputable evidence for or against the cause-and-effect hypothesis" (Hill, 1965, p. 299).

A second problem is *requiring that all causal inferences be based on statistically-significant data*. However, Hill (1965, p. 299; see U.S. EPA, 2005, pp. 2–12, 2–13) warns that "none [of the BHG] can be required as a *sine qua non*," a necessary or sufficient condition, and "no formal tests of significance" answer all cause-effect questions. Often significance tests "contribute nothing to the 'proof' … [and] are totally unnecessary … The glitter of the *t* table diverts attention from the inadequacies of the fare … Often we [incorrectly] deduce 'no difference' from 'no significant difference'" (Hill, 1965, pp. 299–300).

Hill's criticisms warn against a variant of the logical fallacy known as "appeal to ignorance," interpreting the null hypothesis as true/false, based on whether results are/are not statistically significant. Yet, given different circumstances of

complexity, uncertainty, and empirical underdetermination, no BHG could be a necessary condition for asserting causal conclusions in all causal studies. Besides, requiring all data to be statistically significant would beg the question of what evidence, under what conditions, might be most relevant in a given case (Shrader-Frechette, 2011; Swaen and van Amelsvoort, 2009).

In observational research, requiring statistical significance is wrong because significance tests allow reliable inferences only when experimental data are randomized and controlled. Otherwise, sampling bias, selection bias, confounding, and other problems could allow unreliable but statistically-significant results. Because statistical-significance tests presuppose linearity, they likewise have little role in assessing causal effects that are nonlinear, complex, or interactive (Greenland, 1990; Shrader-Frechette, 2011).

Even for experimental research, requiring statistical significance can be suspect because there is no non-arbitrary rationale for choosing $p = .05$; any chosen confidence level may not include the null point, and other p values might show the data are consistent with an effect. Fisher himself recommended that experimenters present all observed p values, not just those meeting the designated value (Shrader-Frechette 2011).

Hill and the U.S. EPA likewise warn against a third BHG error, *ignoring semi-experimental evidence. This error involves interpreting experimental manipulation too narrowly, by discounting important observational evidence*, despite the fact that both "experimental and semi-experimental" manipulation provides the "strongest support" for causal claims (Hill, 1965, pp. 298–299, U.S. EPA 2005, pp. 2–11, 2011, p. 28). For at least 6 reasons, Hill accepts good observational evidence for causality. He argues that sometimes observational studies do as follows:

- provide "semi-experimental" results that explicitly meet the experimental-manipulation guideline (Hill, 1965, p. 298);
- "change" the purported cause, so as to assess any change in purported effect, as when increased cotton dust causes increased harm to mill workers (Hill, 1965, pp. 298–299; U.S. EPA, 2005, pp. 2–15);
- are reliable in controlling for known confounders and using "prospective" cohorts (Hill, 1965, pp. 297);
- are "so clear cut … the contrast between respiratory and nonrespiratory causes of illness so specific, that no formal [experimental] tests could really contribute anything of value to the argument," as cotton-dust-worker cases illustrate (Hill, 1965, p. 299);
- are so compelling that not following them would "weaken our capacity to interpret data and to take reasonable decisions" (Hill, 1965, p. 300); and

- must play a dominant role, given ethics constraints and that only "occasionally" are human-experimental data available: "More often than not, we are dependent upon our observations" for causal inferences (Hill, 1965, pp. 298, 295).

Interestingly, the prominent BHG errors just discussed (trimming the guidelines, requiring statistically-significant data, and ignoring semi-experimental evidence), against which Hill and the U.S. EPA warned, sometimes may undercut the most-used BHG. For instance, in cancer-causality decisions, the International Agency for Research on Cancer (IARC) appears to rely most heavily on the three BHG of experimental manipulation, strength, and consistency (Swaen and van Amelsvoort, 2009), all of which are undercut by the three errors. For instance, erroneously requiring statistically-significant results "trims the data" available for a positive causal inference, both by narrowing the BHG, "experimental manipulation" so as to exclude observational and cohort studies that control for known confounders, and by reducing the "strength" and "consistency" BHG to purely mathematical norms. Contrary to such data-trimming, Hill (1965, p. 296) notes that the same observational results, obtained from "a wide variety of situations and techniques," can provide convincing evidence of causality.

Guideline-trimming errors

How well do the main scientists who use BHG to assess the pesticide-harm hypothesis avoid the errors of trimming guidelines, requiring statistically-significant data, and ignoring semi-experimental evidence? The next three sections argue the Burns-Li-Mink studies, which fail to accept this hypothesis, appear to make these errors, while hypothesis supporters—the EHP-award winners—do not.

In assessing the pesticide-harm hypothesis, all three EHP-award studies explicitly or implicitly use all 9 BHG. They employ strength (Bouchard et al. 2011, pp. 1189–1194; Engel et al., 2011, pp. 1182, 1185–1186; Rauh et al., 2011, pp. 1196–1200); consistency (Bouchard et al., 2011, pp. 1189, 1192–1193; Engel et al., 2011, pp. 1182–1187; Rauh et al., 2011, pp. 1196, 1200); specificity (Bouchard et al., 2011, pp. 1190–1194; Engel et al., 2011, pp. 1182–1184, 1187; Rauh et al., 2011, pp. 1196, 1200); temporality (Bouchard et al., 2011, pp. 1190–1191, 1194; Engel et al., 2011, pp. 1182–1187; Rauh et al., 2011, pp. 1196–1200); biological gradient (Bouchard et al., 2011, p. 1191; Engel et al., 2011, pp. 1186–1187; Rauh et al., 2011, p. 1199); plausibility (Bouchard et al., 2011, p. 1189; Engel et al., 2011, p. 1186; Rauh et al., 2011, pp. 1196, 1200); coherence (Bouchard et al., 2011, pp. 1192–1193; Engel et al., 2011, p. 1187; Rauh et al., 2011, pp. 1196, 1200); experimental manipulation (Bouchard et al., 2011, pp. 1189–1192, 1194; Engel et al., 2011, pp. 1182, 1184, 1187; Rauh et al., 2011, pp. 1196–1198, 1200); and analogy (Bouchard et al., 2011, pp. 1189, 1192; Engel et al., 2011, pp. 1182, 1184;

Rauh et al., 2011, p. 1196). None of these studies errs by trimming guidelines, requiring statistically-significant data, or ignoring semi-experimental data.

However, all major post-2011 studies that fail to accept the pesticide-harm hypothesis commit all three errors and ignore the BHG, experimental manipulation. First, they trim the BHG, use half of them, and thereby perform incomplete causal analyses. Burns et al. (2013, p. 129) use only 4 BHG: strength-of-association, consistency, biological gradient, and plausibility. Li et al. (2012, p. 116) use only 4 BHG: strength-of-association, consistency, biological gradient (narrowed to "dose response"), and temporality. Mink et al. (2012, p. 281) employ these 4, plus biological plausibility.

For instance, Burns et al. (2013), Li et al. (2012), and Mink, Kimmel, and Li (2012) ignore *analogy*, how similar modes-of-action, chemical/metabolite structures, or population patterns support the pesticide-harm hypothesis. They ignore similar patterns of fetal organophosphate exposures in different populations, different exposure effects on IQ, and different human/animal modes-of-action (Bouchard et al., 2011; Bradman et al., 2005; Eskenazi, Bradman, and Castorina, 1999). They ignore analogous human/animal results, e.g., mechanisms of neurodevelopmental harm (Rauh et al., 2011, pp. 1196–1200); analogous cognitive-deficit results from pesticides/other toxicants (Rauh et al., 2006, p. e1855); and analogous, animal/human, pesticide-associated biomarkers, such as certain genotypes (e.g., Engel et al., 2011, p. 1182; Wolff et al., 2007). Yet analogous evidence can support the hypothesis and suggest similar pesticide modes-of-action/structural analogues among different populations (U.S. EPA, 2005, pp. 2–15).

Similarly, by ignoring the BHG *coherence* of data with "other lines of evidence," Burns et al. (2013), Li et al. (2012), and Mink, Kimmel, and Li (2012) ignore support for the pesticide-harm hypothesis from convergent lines of evidence and different studies, including similar mechanisms, e.g., acetylcholinesterase inhibition and paraoxonase detoxification (Bouchard et al., 2011; Engel et al., 2011; Rauh et al., 2011; Wolff et al., 2007) and similar animal (Bouchard et al., 2011; Engel et al., 2011; London et al., 2012; Rauh et al., 2011) and human results (e.g., Bouchard et al., 2011; Engel et al., 2011; Julvez and Grandjean, 2009; London et al., 2012; Rauh et al., 2011). In fact, National Institute of Health program directors Gray and Lawler (2011, p. A328) specifically praise the three 2011 EHP hypothesis-supporting studies for meeting the coherence BHG, for having both "strength in numbers" and "rigorous study design," including prospective birth cohorts.

Similarly, by ignoring BHG *temporality*, Burns et al. (2013) disregard scientific findings showing long-term cognitive problems from pesticide exposure. Decades-long studies of birth cohorts, associating measured pesticide/metabolite levels during gestation with behavioral endpoints throughout childhood/adolescence, document prenatal-exposure effects 1 year (Engel et al., 2011), 2 years (Engel et al., 2011; Eskenazi et al., 2007), 3 years (Rauh et al., 2006; Wolff et al., 2007), and 5–9 years later (Bouchard et al.,

2011; Engel et al., 2011; Marks et al., 2010). Because the three classic EHP studies (Bouchard et al., 2011; Engel et al., 2011; Rauh et al., 2011) avoid guideline trimming, they appear superior--all other things being equal--to the three main studies challenging the pesticide-harm hypothesis (Burns et al., 2013; Li et al., 2012; Mink, Kimmel, and Li, 2012).

Requiring statistically-significant data

Although the main proponents of the pesticide-harm hypothesis use a variety of data, some statistically significant and some not, the key scientists who challenge this hypothesis require statistically-significant data and reduce most BHG to statistical significance. For instance, using the BHG *consistency*, Burns et al. (2013, p. 130; see Li et al., 2012, p. 132, Mink et al., 2012), pp. 303–306) rejects "positive" and "consistent" results that are not statistically significant.

To assess the BHG *strength-of-association*, Burns et al. (2013, p. 130) and Li et al. (2012, p. 132) reject high relative risks that are not statistically significant, and Burns et al. (2013), Li et al. (2012), and Mink, Kimmel, and Li (2012) all explicitly reject observational/epidemiological results merely because they are not-statistically-significant. Although Burns et al. (2013, pp. 219, 258–259) present their articles as independent of Li et al. (2012) and Mink, Kimmel, and Li (2012), they all misrepresent/redefine BHG in terms of statistical significance.

The main problem is not that Burns et al. (2013), Li et al. (2012), and Mink, Kimmel, and Li (2012) use statistical significance at the .05 level, but that they make it a *necessary* condition for their judgments of positive, consistent, or strong causal results, regardless of how well that evidence satisfies the 9 BHG. As already argued, causal assessment requires many complex judgments, e.g., ranking better-designed, better-controlled studies higher, not automatically rejecting observational results as not statistically significant.

Of course, Burns et al. (2013), Li et al. (2012), and Mink, Kimmel, and Li (2012) are not alone in requiring statistical significance. The error is "ubiquitous" in epidemiological/observational studies (Rothman, 2002, p. 126), and courts also often wrongly require toxic-tort plaintiffs to provide statistically-significant evidence of harm (Cranor, 2006, p. 227). Many scientific journals likewise demand statistically-significant evidence for causal conclusions (e.g., Barrett et al., 2005). Some surveys show 63% of journal articles presuppose that statistical nonsignificance establishes no effect, and that statistical significance is necessary for establishing causality (Fidler et al., 2006).

For Burns et al. (2013), Li et al. (2012), and Mink, Kimmel, and Li (2012), to require statistically-significant data is especially serious, given their reducing/redefining BHG, in ways Hill explicitly rejects. They also beg the question against the pesticide-harm hypothesis, for instance, by ignoring animal-experiment studies that provide statistically-significant support for the

pesticide-harm hypothesis, yet assessing only nonrandomized, observational studies, for which statistical significance is irrelevant (Burns et al., 2013, p. 127). All three studies also confuse the absence of evidence for an effect with evidence for the absence of an effect; they assume a not-statistically-significant result is evidence of no effect (Greenland, 2011; Shrader-Frechette, 2011). Because of wrongly requiring statistically-significant data, the three main studies that criticize the pesticide-harm hypothesis appear inferior, all other things being equal, to the classic EHP studies that support the hypothesis and avoid this error.

Ignoring semi-experimental evidence

The main studies that criticize the pesticide-harm hypothesis likewise ignore semiexperimental and experimental evidence, the "strongest support" for causal hypotheses (Hill, 1965, pp. 298–299; U.S. EPA, 2005, pp. 2–14ff), while the main EHP hypothesis-supporting research (e.g., Bouchard et al., 2011; Bradman et al., 2005; Engel et al., 2011; Eskenazi, Bradman, and Castorina, 1999, 2007; Marks et al., 2010; Rauh et al., 2006, 2011; Wolff et al., 2007) avoid this error. As shown, the three EHP studies use all 9 BHG. They understand that good observational studies can substantially satisfy the experimental manipulation BHG if they show that a "change in exposure brings about a change in disease frequency" (U.S. EPA, 2005, pp. 2–14ff).

In at least 7 ways, (1)–(7), the Berkeley/Columbia/Mt. Sinai or EHP studies partly satisfy the experimental manipulation BHG by showing that, as pesticide exposure increases, so does frequency of neurodevelopmental problems. That is, they strengthen the semi-experimental nature of their observational studies by using 7 techniques, including controlling for known confounders; they use (1) *prospective*, (2) *long-term*, and (3) *birth cohorts* that are the human-exposure "gold standard," and they assess (4) geographically *diverse* populations, respectively, in California; New York; and China, Poland, and New York. Consequently, these hypothesis-supporting studies have less recall bias (given regular-interval data collection over 16–17 years) and less sampling bias (because they track diverse groups, without obvious initial effects, and compare them to those without any effects).

The 11 Berkeley/Columbia/Mt. Sinai cohorts are also stronger because they are (5) *methodologically diverse*, e.g., measuring pesticides in umbilical-cord blood (Rauh et al., 2011) or pesticide metabolites in maternal-urine samples during pregnancy (Bouchard et al., 2011; Engel et al., 2011). Because they are (6) *large*, each with roughly 500 people and a total of 6,000 (Bouchard et al., 2011; Engel et al., 2011; Rauh et al., 2011), they reduce chances of false-negative errors. Finally, the 11 cohorts are methodologically strong in showing pesticide-caused neurodevelopmental harm because all studies (7) *control for known confounders*, such as other non-neurodevelopmental endpoints

(including asthma, behavioral disorders, cancer, chromosomal abnormalities, genetic defects, and obesity); other exposures (such as allergens, bisphenol-A, cigarette smoke, ether, lead, mercury, mold, naphalene, phthalates, polybrominated diphenyl ethers, and polycyclic aromatic hydrocarbons); and other influences on subjects (such as maternal alcohol use, breastfeeding, education, genetics, housing type, and race).

Because Bouchard et al. (2011), Engel et al. (2011), and Rauh et al. (2011) exhibit preceding strength (7), *control for confounders*, they satisfy much of the BHG of experimental manipulation. They use excellent experimental design to show that neurodevelopmental damage increases as pesticide exposure increases. They also allow for multiple "natural experiments" that examine different doses/effects. Both reasons go a long way toward satisfying the experimental-manipulation BHG, though epidemiological research can never fully meet it (Phillips and Goodman, 2006).

However, Burns et al. (2013), Li et al. (2012), and Mink, Kimmel, and Li (2012) completely ignore experimental manipulation, although, as argued earlier, Hill says it is the dominant/strongest BHG. They also fail to assess the semi-experimental strengths of the 11 long-term, prospective cohort studies that support the pesticide-harm hypothesis. Hence, because the three EHP-award-winning, pesticide-hypothesis-supporting studies (Bouchard et al., 2011; Engel et al., 2011; Rauh et al., 2011) meet the experimental manipulation BHG as Hill defines it, they appear superior—all other things being equal—to the three main studies (Burns et al., 2013; Li et al., 2012; Mink, Kimmel, and Li, 2012) that challenge this hypothesis, that ignore this BHG, and that fail to consider what Hill (1965, p. 298) calls "semi-experimental" evidence.

If the preceding analysis is correct, the EHP-prize-winning studies may support the pesticide-harm hypothesis, at least in part because they avoid three major BHG errors—trimming guidelines, requiring statistically-significant data, and ignoring semi-experimental evidence. However, the Burns et al. (2013), Li et al. (2012), and Mink, Kimmel, and Li (2012) studies fail to accept this hypothesis, apparently in part because they commit all three errors. All other things being equal, accepting (rather than not accepting) the pesticide-harm hypothesis appears more plausible.

Explaining obvious methodological errors

One question is why three different groups of scientists would publish studies that make many of the exact BHG causal errors, against which Hill explicitly warned, then take a position critical of the developing scientific consensus in favor of the pesticide-harm hypothesis. A possible explanation is financial conflicts of interest (NRC, 2004), "pervasive" in industry-funded research (Krimsky, 2004; Needleman et al., 2005; Shrader-Frechette, 2007), especially environmental-health research (e.g., Grandjean and Ozonoff, 2013; Michaels, 2008).

All major, post-2011 studies discussed here, critical of the pesticide-harm hypothesis, are funded or written by employees of Dow Chemical Company, a major pesticide manufacturer (Burns et al., 2013, p. 127; Li et al., 2012, p. 109; Mink, Kimmel, and Li, 2012, p. 281). And more often than not, knowing a study's funders predicts its conclusions (e.g., Barnes and Bero, 1998; Bekelman, Li, and Gross, 2003; Krimsky, 2004).

Conclusions

Those who claim to use BHG and U.S. EPA weight-of-evidence causal guidelines, yet systematically trim/misuse them, as Burns et al. (2013), Li et al. (2012), and Mink, Kimmel, and Li (2012) appear to do, cannot validly claim inadequate grounds for accepting the pesticide-harm hypothesis. All other things being equal, accepting this hypothesis appears more defensible than not accepting it.

Funding

This research was made possible by grants SES-0724781 and SES-82-05112 to K.S-F. (principal investigator) from the U.S. National Science Foundation (NSF). The contents of this article are solely the responsibility of the authors, not NSF.

References

Barnes, D. E., and L. A. Bero. 1998. Why review articles on the health effects of passive smoking reach different conclusions. *JAMA* 279:1566–70. doi:10.1001/jama.279.19.1566.

Barrett, K. A., C. L. Funk, and F. L. Macrina. 2005. Awareness of publication guidelines and the responsible conduct of research. *Accountability in Research* 12:193–206. doi:10.1080/08989620500217321.

Bekelman, J. E., Y. Li, and C. P. Gross. 2003. Scope and impact of financial conflicts of interest in biomedical research: A systematic review. *JAMA* 289:454–65. doi:10.1001/jama.289.4.454.

Bouchard, M. F., J. Chevrier, K. G. Harley, K. Kogut, M. Vedar, N. Calderon, C. Trujillo, C. Johnson, A. Bradman, D. B. Barr, and B. Eskenazi. 2011. Prenatal exposure to organophosphate pesticides and IQ in 7-year old children. *Environmental Health Perspectives* 119:1189–95. doi:10.1289/ehp.1003185.

Bradman, A., B. Eskenazi, D. B. Barr, R. Bravo, R. Castorina, J. Chevrier, K. Kogut, M. E. Harnly, and T. E. McKone. 2005. Organophosphate urinary metabolite levels during pregnancy and after delivery. *Environmental Health Perspectives* 113:1802–07. doi:10.1289/ehp.7894.

Burns, C. J., L. J. McIntosh, P. J. Mink, A. M. Jurek, and A. A. Li. 2013. Pesticide exposure and neurodevelopmental outcomes. *Journal of Toxicology and Environmental Health. Part B, Critical Reviews* 16:127–283. doi:10.1080/10937404.2013.783383.

Cranor, C. F. 2006. *Toxic torts*. Cambridge, UK: Cambridge University Press.

Engel, S. M., J. Wetmur, J. Chen, C. Zhu, D. Boyd Barr, R. L. Canfield, and M. S. Wolff. 2011. Prenatal exposure to organophosphates, paraoxonase 1, and cognitive development in childhood. *Environmental Health Perspectives* 119:1182–88. doi:10.1289/ehp.1003183.

Eskenazi, B., A. Bradman, and R. Castorina. 1999. Exposures of children to organophosphate pesticides and their potential adverse health effects. *Environmental Health Perspectives* 107:409–19. doi:10.1289/ehp.99107s3409.

Eskenazi, B., A. R. Marks, A. Bradman, K. Harley, D. B. Barr, C. Johnson, N. Morga, and N. P. Jewell. 2007. Organophosphate pesticide exposure and neurodevelopment in young Mexican-American children. *Environmental Health Perspectives* 115:792–98. doi:10.1289/ehp.9828.

Fidler, F., M. A. Burgman, G. Cumming, R. Buttrose, and N. Thomason. 2006. Impact of criticism of null-hypothesis significance testing on statistical reporting practices in conservation biology. *Conservation Biology* 20:1539–44. doi:10.1111/j.1523-1739.2006.00525.x.

Grandjean, P., and P. J. Landrigan. 2014. Neurobehavioural effects of developmental toxicity. *The Lancet Neurology* 13:330–38. doi:10.1016/S1474-4422(13)70278-3.

Grandjean, P., and D. Ozonoff. 2013. Transparency and translation of science in a modern world. *Environmental Health* 12:70. doi:10.1186/1476-069X-12-70.

Gray, K. and C. P. Lawler. 2011. Strength in numbers: Three separate studies link in utero organophosphate pesticide exposure and cognitive development. *Environmental Health Perspectives* 119:a328–a329.

Greenland, S. 1990. Randomization, statistics, and causal inference. *Epidemiology* 1:421–29. doi:10.1097/00001648-199011000-00003.

Greenland, S. 2011. Null misinterpretation in statistical testing and its impact on health risk assessment. *Preventive Medicine* 53:225–28. doi:10.1016/j.ypmed.2011.08.010.

Hill, A. B. 1965. The environment and disease? *Proceedings of the Royal Society of Medicine* 58:295–300.

Julvez, J., and P. Grandjean. 2009. Neurodevelopmental toxicity risks due to occupational exposure to industrial chemicals during pregnancy. *Industrial Health* 47:459–68. doi:10.2486/indhealth.47.459.

Krimsky, S. 2004. Introduction to special issue on conflict of interest in science. *Accountability in Research* 11:79–81. doi:10.1080/03050620490512214.

Landrigan, P. J., L. Lambertini, and L. S. Birnbaum. 2012. A research strategy to discover the environmental causes of autism and neurodevelopmental disabilities. *Environmental Health Perspectives* 120:a258–a260. doi:10.1289/ehp.1104285.

Li, A. A., K. A. Lowe, L. J. McIntosh, and P. J. Mink. 2012. Evaluation of epidemiology and animal data for risk assessment: Chlorpyrifos developmental neurobehavioral outcomes. *Journal of Toxicology and Environmental Health. Part B, Critical Reviews* 15:109–84. doi:10.1080/10937404.2012.645142.

London, L., C. Beseler, M. F. Bouchard, D. C. Bellinger, C. Colosio, P. Grandjean, R. Harari, T. Kootbodien, H. Kromhout, F. Little, T. Meijster, A. Moretto, D. S. Rohlman, and L. Stallones. 2012. Neurobehavioral and neurodevelopmental effects of pesticide exposures. *Neurotoxicology* 33:887–96. doi:10.1016/j.neuro.2012.01.004.

Marks, A. R., K. Harley, A. Bradman, K. Kogut, D. B. Barr, C. Johnson, N. Calderon, and B. Eskenazi. 2010. Organophosphate pesticide exposure and attention in young Mexican-American children. *Environmental Health Perspectives* 118:1768–74. doi:10.1289/ehp.1002056.

Michaels, D. 2008. *Doubt is their product*. Oxford, New York: Oxford University Press.

Mink, P. J., C. A. Kimmel, and A. A. Li. 2012. Potential effects of chlorpyrifos on fetal growth outcomes. *Journal of Toxicology and Environmental Health. Part B, Critical Reviews* 15:281–316. doi:10.1080/10937404.2012.672150.

Phillips, C. V. and K. J. Goodman. 2006. Causal criteria and counterfactuals. *Emerging Themes in Epidemiology* 3:5. doi:10.1186/1742-7622-3-5.

Needleman, H. L., J. R. Reigart, P. Landrigan, J. Sass, and C. Bearer. 2005. Benefits and risks of pesticide testing on humans. *Environmental Health Perspectives* 113:A804–A805. doi:10.1289/ehp.113-a804.

NRC (National Research Council). 2004. *The ethical conduct of clinical research involving children*. Washington, D.C.: National Academies Press.

Rauh, V., S. Arunajadai, M. Horton, F. Perera, L. Hoepner, D. B. Barr, and R. Whyatt. 2011. Seven-year neurodevelopmental scores and prenatal exposure to chlorpyrifos. *Environ Health Perpect* 119:1196–201. doi:10.1289/ehp.1003160.

Rauh, V. A., R. Garfinkel, F. P. Perera, H. F. Andrews, L. Hoepner, D. B. Barr, R. Whitehead, D. Tang, and R. W. Whyatt. 2006. Impact of prenatal chlorpyrifos exposure on neurodevelopment in the first 3 years of life. *Pediatrics* 118:e1845–e1859. doi:10.1542/peds.2006-0338.

Rothman, K. J. 2002. *Epidemiology*. New York: Oxford University Press.

Shrader-Frechette, K. 2007. EPA's 2006 human-subjects rule for pesticide experiments. *Accountability in Research* 14:211–54. doi:10.1080/08989620701670161.

Shrader-Frechette, K. 2011. Randomization and rules for causal inferences in biology. *Biological Theory* 6:154–61. doi:10.1007/s13752-012-0021-y.

Swaen, G., and L. van Amelsvoort. 2009. A weight of evidence approach to causal inference. *Journal of Clinical Epidemiology* 62:270–77. doi:10.1016/j.jclinepi.2008.06.013.

U.S EPA (Environmental Protection Agency). 2005. *Guidelines For Carcinogen Risk Assessment*. Washington, D.C.: Risk Assessment Forum, U.S. Environmental Protection Agency. http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OA-2007-0679-0001 (accessed December 22, 2015).

U.S. EPA (Environmental Protection Agency). 2010. *Framework for incorporating human Epidemiologic & Incident data in health risk assessment (DRAFT)*. Washington, D.C: U.S. Environmental Protection Agency. http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPP-2010-0125-0012 (accessed December 22, 2015).

U.S. EPA (Environmental Protection Agency). 2011. *Weight-of-evidence*. Washington, D.C: U.S. Environmental Protection Agency. http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPPT-2010-0877-0021 (accessed December 22, 2015).

Wolff, M. S., S. Engel, G. Berkowitz, S. Teitelbaum, J. Siskind, D. B. Barr, and J. Wetmur. 2007. Prenatal pesticide and PCB exposures and birth outcomes. *Pediatric Research* 61:243–50. doi:10.1203/pdr.0b013e31802d77f0.