Statistics for Biology and Health

Series Editors:

For other titles published in this series, go to
http://www.springer.com/series/2848

Timothy L. Lash • Matthew P. Fox
Aliza K. Fink

Applying Quantitative Bias Analysis to Epidemiologic Data

Springer

Timothy L. Lash
Boston University
School of Public Health
715 Albany St.
Boston, MA 02118, USA

Matthew P. Fox
Boston University
School of Public Health
715 Albany St.
Boston, MA 02118, USA

Aliza K. Fink
Boston University
School of Public Health
715 Albany St.
Boston, MA 02118, USA

ISBN: 978-0-387-87960-4 e-ISBN: 978-0-387-87959-8
DOI: 10.1007/978-0-387-87959-8
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2009921156

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Bias analysis quantifies the influence of systematic error on an epidemiology study's estimate of association. The fundamental methods of bias analysis in epidemiology have been well described for decades, yet are seldom applied in published presentations of epidemiologic research. More recent advances in bias analysis, such as probabilistic bias analysis, appear even more rarely. We suspect that there are both supply-side and demand-side explanations for the scarcity of bias analysis. On the demand side, journal reviewers and editors seldom request that authors address systematic error aside from listing them as limitations of their particular study. This listing is often accompanied by explanations for why the limitations should not pose much concern. On the supply side, methods for bias analysis receive little attention in most epidemiology curriculums, are often scattered throughout textbooks or absent from them altogether, and cannot be implemented easily using standard statistical computing software. Our objective in this text is to reduce these supply-side barriers, with the hope that demand for quantitative bias analysis will follow.

We began this project after writing a few bias analysis papers in epidemiology journals. The American College of Epidemiology invited us to prepare a day-long workshop on the methods of bias analysis, which became an outline, which in turn became this text. The text begins with a description of the need for bias analysis and continues with two chapters on the legwork that must be done at the outset of a research project to plan and implement a bias analysis. This introductory section is followed by three chapters that explain simple bias analysis methods to address each of the fundamental threats to the validity of an epidemiology study's estimate of association: selection bias, classification errors, and uncontrolled confounding. We then extend the bias analysis methods from these early chapters to multidimensional, probabilistic, and ultimately multiple bias analysis methods. The book concludes with a chapter on the presentation and interpretation of the results of a bias analysis.

Readers might use the text as an independent resource to address bias analysis as they conduct epidemiologic research, as a secondary text in a class on epidemiologic methods, or as the central text in an advanced class on data analysis in epidemiologic research that focuses on bias analysis. We hope that students will find, as we have, that once they have completed one bias analysis, it becomes hard to imagine

analyzing epidemiologic data without it. We have aimed the text at readers who have some familiarity with epidemiologic research and intermediate data analysis skills. For those without those skills, we suggest a comprehensive methods text, such as *Modern Epidemiology*, which can be used in conjunction with this text to provide a foundation in epidemiologic terminology, study design, and data analysis. Readers with advanced skills, particularly statistical skills, might yearn for a fully Bayesian treatment of the topic of bias analysis. Our approach is intentionally more fundamental, in the hope that a wider audience of epidemiologists and data analysts will adopt bias analysis methods if they do not have to simultaneously confront the barriers (real or perceived) of Bayesian statistics.

An important adjunct resource for this textbook is the suite of freely available spreadsheets and software available for download at https://sites.google.com/site/biasanalysis/. (N.B. The credit for development of these tools goes solely to Matthew Fox, whose perseverance and vision to enable bias analysis by these techniques has added substantial value to the text.) We encourage readers to download the software, follow the examples in the text, and then modify the fields to implement their own bias analysis. We would be delighted to hear from anyone who improves the tools or detects an error, and we will post revised tools as they become available. Likewise, we welcome comments, criticisms, and errata regarding the text from readers and will maintain a log of this feedback on the aforementioned web site.

In closing, we thank our friends and colleagues who contributed to the text directly or indirectly. We appreciate Charles Poole's suggestion to the American College of Epidemiology that a course on bias analysis would be of value, and we appreciate John Acquavella's decision to accept that suggestion on behalf of the college. Sander Greenland participated in the development and presentation of the American College of Epidemiology workshops, and has been instrumental in improving the methods of bias analysis. We are very grateful for his input and dedication to the topic. We also thank our colleagues who have a particular interest in bias analysis methods; they have challenged us to develop our ideas and to communicate them clearly. We cannot list them all, so acknowledge especially Charles Poole, Carl Phillips, George Maldonado, Anne Jurek, Ken Rothman, Rebecca Silliman, Soe Soe Thwin, Dan Brooks, and Steve Cole. We also acknowledge the important contribution of three anonymous reviewers recruited by our patient publisher; perhaps some have already been named above.

Funding for this project was made possible by grant 1 G13 LM008530 from the National Library of Medicine, NIH, DHHS. The views expressed in any written publication, or other media, do not necessarily reflect the official policies of the Department of Health and Human Services; nor does mention by trade names, commercial practices, or organizations imply endorsement by the U.S. Government. Sections of Chapter 1 first appeared in: Lash TL. Heuristic thinking and inference from observational epidemiology. *Epidemiology* 2007;18(1):67–72 and are used here with permission. Examples of misclassified data in Chapters 6, 7, and 8 are used with kind permission from Springer Science+Business Media and first appeared in: Fink AK, Lash TL. A null association between smoking during

pregnancy and breast cancer using Massachusetts registry data. *Cancer Causes and Control* 2003;14(5):497–503. The example of multidimensional bias analysis involving colorectal cancer mortality and 5-fluorouracil treatment is used with permission and first appeared in: Sundararajan V, Mitra N, Jacobson JS, Grann VR, Heitjan DF, Neugut AI. Survival associated with 5-fluorouracil-based adjuvant chemotherapy among elderly patients with node-positive colon cancer. *Annals of Internal Medicine* 2002;136(5):349–357.

Boston, MA

Timothy L. Lash
Matthew P. Fox
Aliza K. Fink

Contents

Chapter 1
Introduction, Objectives, and an Alternative

Introduction

Nonrandomized Epidemiologic Research

The results of nonrandomized epidemiologic investigations have a direct impact on all aspects of health interventions. Studies of social, environmental, behavioral, and molecular risk factors associated with the incidence of particular diseases lead to primary public health interventions aimed at preventing the disease from occurring. Few studies of etiologic relations allow for the exposure to be assigned by randomization because of ethical constraints; participants cannot be randomized ethically to an exposure that might cause harm. Secondary interventions aim to reduce the disease burden by detecting disease before symptoms manifest, so that treatments can more effectively cure the disease or reduce its morbidity. While many studies of disease-screening programs are conducted by randomized designs, some have been conducted using nonrandomized designs (Weiss, 1994). In addition, the efficacy of screening programs established by randomized designs is often compared with its effectiveness measured by nonrandomized designs (Weiss, 1994), and history of screening can be an important confounder of etiologic relations (Weiss, 2003). Tertiary interventions, or medical interventions, aim to reduce the disease burden by curing the disease or by reducing its morbidity. Ideally, the efficacy of medical interventions is established by randomized study designs. However, such designs are sometimes unethical when patients cannot be assigned to a valid comparison group. For example, patients cannot be assigned to receive a placebo or to receive no therapy when there are accepted medical interventions available. When such a comparison group is required, nonrandomized designs are the only alternative. The medical literature contains a continuous and vigorous discussion about the advantages and disadvantages of nonrandomized versus randomized controlled trial evidence (Barton, 2000; Ioannidis et al., 2001a, b) and about the role of both in evidence-based medicine. Randomized controlled trials and nonrandomized studies have complementary roles (Sorensen et al., 2006), particularly when external validity, feasibility, and ethical concerns are paramount. Furthermore, nonrandomized designs

T.L. Lash et al. *Applying Quantitative Bias Analysis to Epidemiologic Data*,
DOI: 10.1007/978-0-387-87959-8_1, © Springer Science+Business Media, LLC 2009

provide a measure of the effectiveness of therapies – for which efficacy has been established by randomized designs – in clinical practice settings that involve patients with characteristics that differ from the clinical trial subjects (e.g., the elderly or other underserved subpopulations). Thus, nonrandomized epidemiologic research contributes to the knowledge base for disease prevention, early detection, and treatment.

The Treatment of Uncertainty in Nonrandomized Research

If the objective of epidemiologic research is to obtain a valid, precise, and generalizable estimate of the effect of an exposure on the occurrence of an outcome (e.g., disease), then investigators have a twofold obligation. First, they must design their investigations to enhance the precision and validity of the effect estimate that they obtain. Second, recognizing that no study is perfect, they must inform stakeholders (collaborators, colleagues, and consumers of their research findings) how near the precision and validity objectives they believe their estimate of effect might be.

To enhance the precision of an effect estimate (i.e., to reduce random error), epidemiologists design their studies to gather as much information as possible (Rothman et al., 2008a), apportion the information efficiently among the strata of variables that affect the outcome (Rothman et al., 2008a), and undertake precision-enhancing analyses such as pooling (Greenland and Rothman, 2008) and regression (Greenland, 2008). The methods to enhance a study's validity and the precision of its estimate of effect are well-described in epidemiologic textbooks such as *Modern Epidemiology* (Rothman et al., 2008b). Even with an efficient design and analysis, epidemiologists customarily present a quantitative assessment of the remaining random error about an effect estimate. Although there has been considerable (Thompson, 1987a, b; Poole, 1987, b) and continuing (The Editors, 2001; Weinberg, 2001; Gigerenzer, 2004) debate about methods of describing random error, a consensus has emerged in favor of the frequentist confidence interval (Poole, 2001).

To enhance the validity of an effect estimate (i.e., to reduce systematic error), epidemiologists design their studies to assure comparability of the exposed and unexposed groups (Greenland and Robins, 1986), reduce differential selection forces (Miettinen, 1985; Wacholder et al., 1992), and control measurement error by obtaining accurate information or forcing the direction of its expected error to be predictable (Greenland, 1980; Brenner and Savitz, 1990). When the validity might be compromised by confounding after implementation of the design, epidemiologists employ analytic techniques such as stratification (Greenland and Rothman, 2008) or regression (Greenland, 2008) to improve the validity of the effect estimate. Analytic corrections for selection forces or measurement error are seldom seen. Quantitative assessments of the remaining systematic error about an effect estimate are even more rare.

Thus, the quantitative assessment of the error about an effect estimate usually reflects only the residual random error. Much has been written (Poole, 1987a, b, 2001; Gigerenzer, 2004; Lang et al., 1998) and many examples proffered (Rothman, 1999; Lash, 1998) about the abuses made of these quantitative assessments of random error.

The near complete absence of quantitative assessments of residual systematic error in published epidemiologic research has received much less attention. Several reasons likely explain this inattention. First, existing custom does not expect a quantitative assessment of the systematic error about an effect estimate. For example, the uniform requirements for manuscripts submitted to biomedical journals instructs authors to "quantify findings and present them with appropriate indicators of measurement error or uncertainty (such as confidence intervals)," which measure only residual random error. With no demand to drive development and no habit to breed familiarity, few methods are available to quantify the systematic error about an effect estimate and few epidemiologists are comfortable with the implementation of existing methods. However, recent methods papers published in leading epidemiology (Steenland and Greenland, 2004) and statistics journals (Greenland, 2005) have called for routine training in bias modeling for epidemiology students, so demand for this training will hopefully grow in the near term. Second, the established methods often require presentations of systematic error that are lengthy (Greenland and Lash, 2008), so are too unwieldy to incorporate into data summarization and inference. By comparison, the quantitative assessments of random error require little additional space for presentation of an apparently rigorous measurement of residual random error. Finally, the automated analytic tools often used by epidemiologists provide quantitative assessments of residual random error about effect estimates, but contain no such automated method of assessing residual systematic error.

Objective

The objective of this test is to reduce the aforementioned barriers to regular implementation of quantitative sensitivity analysis. Epidemiologic studies yield effect estimates such as the risk ratio, rate ratio, odds ratio, or risk difference; all of which compare measurements of the occurrence of an outcome in a group with some common characteristic (such as an exposure) with the occurrence of the outcome in a second group with some other common characteristic (such as the absence of exposure). The error accompanying an effect estimate equals the square of its difference from the true effect, and conventionally parses into random error (variance) and systematic error (bias squared). Under this construct, random error is that which approaches zero as the study size increases and systematic error is that which does not. The amount of random error in an effect estimate is measured by its precision, which is usually quantified by *p*-values or confidence intervals that accompany the effect estimate. The amount of systematic error in an effect estimate is measured by its validity, which is seldom quantified. A quantitative assessment of the systematic error about an effect estimate can be made using bias analysis.

In this text, we have collected existing methods of quantitative bias analysis, explained them, illustrated them with examples, and linked them to tools for implementation. The second chapter provides a guide to choosing the method most appropriate for the problem at hand and for making inference from the method's

results. The software tools automate the analysis in familiar software and provide output that reduces the resources required for presentation. Probabilistic bias analysis and multiple biases modeling, for example, yield output that is no more complicated to present and interpret than the conventional point estimate and its associated confidence interval.

While we have compiled a set of methods to address comprehensively the most common threats to a study result's validity (selection bias, information bias, and unmeasured confounding), we have not addressed all possible threats to validity or even all methods to address even these common threats. For example, we have not addressed model misspecification or bias from missing data. We have not addressed empirical methods of bias analysis or Bayesian methods of bias analysis, although these methods are related to many of the methods we do present. The interested reader can find textbooks and journal articles that describe these methods, some of which can be implemented by freely available software that can be downloaded from the internet.

We have not presented these methods for several reasons. First, this text is directed to practicing epidemiologists who are familiar with these threats to validity and who are comfortable with spreadsheets and relatively fundamental SAS® software programming. The alternative methods often require more sophisticated computer programming than required to implement the methods we present. Second, the empirical methods often require assumptions about the accuracy of the data source used to inform the bias analysis, which we believe can seldom be supported. We prefer to recognize that the validation data are often themselves measured with error, and that this error should be incorporated into the bias analysis. The methods we present more readily accommodate this preference. Third, the Bayesian methods are similar to the probabilistic bias analysis methods and probabilistic multiple bias analysis methods we present toward the end of this text. The primary difference is that the Bayesian methods require specification of a prior for the parameter to be estimated (i.e., ordinarily the association between an exposure and an outcome). While we recognize and even agree with this Bayesian approach to data analysis and inference, particularly compared with the inherent frequentist prior that any association is equally likely, this text is not the forum to continue that debate.

An Alternative

As stated earlier, epidemiologic research is an exercise in measurement. Its objective is to obtain a valid and precise estimate of either the occurrence of disease in a population or the effect of an exposure on the occurrence of disease. Conventionally, epidemiologists present their measurements in three parts: a point estimate (e.g., a risk ratio), a frequentist statistical assessment of the uncertainty (e.g., a confidence interval, but also sometimes a *p*-value), and a qualitative description of the threats to the study's validity.

Without randomization of study subjects to exposure groups, point estimates, confidence intervals, and *p*-values lack their correct frequentist interpretations

(Greenland, 1990). Randomization and a hypothesis about the expected allocation of outcomes – such as the null hypothesis – allow one to assign probabilities to the possible outcomes. One can then compare the observed association, or a test statistic related to it, with the probability distribution to estimate the probability of the observed association, or associations more extreme, under the initial hypothesis. This comparison provides an important aid to causal inference (Greenland, 1990) because it provides a probability that the outcome distribution is attributable to chance as opposed to the effects of exposure. The comparison is therefore at the root of frequentist statistical methods and inferences from them. When the exposure is not assigned by randomization, as is the case for nonrandomized epidemiologic research (and for randomized trials with withdrawals or classification errors), the comparison provides a probability that the outcome distribution is attributable to chance as opposed to the combined effects of exposure and systematic errors. Causal inference therefore requires an educated guess about the strength of the systematic errors compared with the strength of the exposure effects.

These educated guesses can be accomplished quantitatively by likelihood methods (Espeland and Hui, 1987), Bayesian methods (Gustafson, 2003), regression calibration (Spiegelman et al., 2000), missing data methods (Little and Rubin, 2002; Robins et al., 1994), or Monte Carlo simulation (Lash and Fink, 2003b; Phillips, 2003; Greenland, 2004) [*see* Greenland (2005) for a review and comparison of methods]. Some of these methods will be described in later chapters. The conventional approach, however, is to make the guess qualitatively by describing the study's limitations. An assessment of the strength of systematic errors, compared with the strength of exposure effects, therefore becomes an exercise in reasoning under uncertainty. Human ability to reason under uncertainty has been well-studied and shown to be susceptible to systematic bias resulting in predictable mistakes. A brief review of this literature, focused on situations analogous to epidemiologic inference, suggests that the qualitative approach will frequently fail to safeguard against tendencies to favor exposure effects over systematic errors as an explanation for observed associations. The aforementioned quantitative methods have the potential to safeguard against these failures.

Heuristics

The Dual-Process Model of Cognition

A substantial literature from the field of cognitive science has demonstrated that humans are frequently biased in their judgments about probabilities and at choosing between alternative explanations for observations (Piattelli-Palmarini, 1994b; Kahneman et al., 1982; Gilovich et al., 2002), such as epidemiologic associations. Some cognitive scientists postulate that the mind uses dual processes to solve problems that require such evaluations or choices (Kahneman and Frederick, 2002; Sloman, 2002). The first system, labeled the "Associative System," uses patterns to draw inferences.

We can think of this system as intuition, although any pejorative connotation of that label should not be applied to the associative system. The second system, labeled the "Rule-Based System," applies a logical structure to a set of variables to draw inferences. We can think of this system as reason, although the label alone should not connote that this system is superior. The Associative System is not necessarily less capable than the Rule-Based System; in fact, skills can migrate from the Rule-Based System to the Associative System with experience. The Associative System is in constant action, while the Rule-Based System is constantly monitoring the Associative System to intervene when necessary. This paradigm ought to be familiar; we have all said at some time "Wait a minute – let me think," by which we do not mean that we have not yet thought, but that we are not satisfied with the solution our Associative System's thought has delivered. After the chance to implement the Rule-Based System, we might say "On second thought, I have changed my mind," by which we mean that the Rule-Based System has overwritten the solution initially delivered by the Associative System.

The process used by the Associative System to reach a solution relies on heuristics. A heuristic reduces the complex problem of assessing probabilities or predicting uncertain values to simpler judgmental operations (Tversky and Kahneman, 1982b). An example of a heuristic often encountered in epidemiologic research is the notion that nondifferential misclassification biases an association toward the null. Heuristics often serve us well because their solutions are correlated with the truth, but they can sometimes lead to systematic and severe errors (Tversky and Kahneman, 1982b). Nondifferential and nondependent misclassification of a dichotomous exposure leads to the expectation that an association will be biased toward the null, but many exceptions exist. For example, any particular association influenced by nondifferential misclassification may not be biased toward the null (Jurek et al., 2005), dependent errors in classification can substantially bias an association away from the null – even if classification errors are nondifferential (Kristensen, 1992), nondifferential misclassification of disease may not lead to any bias in some circumstances (Brenner and Savitz, 1990), and a true association may not provide stronger evidence against the null hypothesis than the observed association based on the misclassified data – even if the observed association is biased toward the null (Gustafson and Greenland, 2006). Application of the misclassification heuristic without deliberation can lead to errors in an estimate of the strength and direction of the bias (Lash and Fink, 2003a), as is true for more general cognitive heuristics (Tversky and Kahneman, 1982b).

Cognitive scientists have identified several classes of general heuristics, three of which are described below because they may be most relevant to causal inference based on nonrandomized epidemiologic results. These heuristics have the following characteristics in common (Piattelli-Palmarini, 1994a). First, the errors in judgments attributable to the heuristic are systematic and directional; that is, they always act in the same way and in the same direction. Second, they are general and nontransferable; that is, all humans are susceptible to the errors and knowledge of how they act does not immunize us against them. Third, they are independent of intelligence and education; that is, experts make the same mistakes as novices,

particularly if the problem is made a little more difficult or moved a small distance outside of their expertise. While studies that have elicited an understanding of these heuristics have most often been conducted in settings that are not very analogous to causal inference using epidemiologic data, one such study has been conducted and its results corresponded to results elicited in the cognitive science setting (Holman et al., 2001). In addition, these heuristics have been shown to affect evidence-based forecasts of medical doctors, meteorologists, attorneys, financiers, and sports prognosticators (Koehler et al., 2002). It seems unlikely that epidemiologists would be immune.

Anchoring and Adjustment

The first heuristic relevant to causal inference based on nonrandomized epidemiologic results is called "anchoring and adjustment" (Tversky and Kahneman, 1982b). When asked to estimate an unknown but familiar quantity, respondents use a heuristic strategy to select (or receive) an anchor, and then adjust outward from that anchor in the direction of the expected true value. Adjustments are typically insufficient. For example, one might be asked to give the year in which George Washington was elected as the first president of the USA (Epley and Gilovich, 2002). Most respondents choose the anchor to be 1776, the year that the USA declared independence. Respondents adjust upward to later years, because they know the US Constitution was not ratified in the same year. The average response equals 1779, and the correct value equals 1788. Why, on average, is the upward adjustment insufficient? The predictably insufficient adjustment may arise because respondents adjust outward from the anchor until their adjusted estimate enters a range they deem plausible. The true value, more often, lies toward the center of the plausible range. When the anchor is below the true value, as in the year that Washington was first elected, the estimate is predictably lower than the true value. Conversely, when the anchor is above the true value, the estimate is predictably higher than the true value. For example, one might be asked to give the temperature at which vodka freezes (Epley and Gilovich, 2002). Most respondents choose the anchor to be 32°F, the temperature at which water freezes. Respondents adjust downward to lower temperatures, because they know alcohol freezes at a lower temperature than water. The average response equals 1.75°F, and the correct value equals −20°F. Importantly, the anchoring and adjustment heuristic operates in the same manner regardless of whether the anchor is self-generated or provided by an external source, so long as the respondent is aware of the anchor and it is on the same scale as the target (Chapman and Johnson, 2002).

How might the anchoring and adjustment heuristic affect inference from nonrandomized epidemiologic results? Consider the point estimate associating an exposure with a disease, derived from a study's results, to be an anchor. Further consider that stakeholders (the investigator, collaborators, readers, and policymakers) may be aware of the direction of an expected bias (e.g., toward the null). Can the stakeholders be expected to adjust the point estimate sufficiently to account for the bias? An understanding of the anchoring and adjustment heuristic suggests that the adjustment will be predictably insufficient. Stakeholders should be expected to adjust the

association to account for the bias only so far as is plausible, which adjustment will, on average, be insufficient.

Overconfidence

The second bias relevant to causal inference based on nonrandomized epidemiologic results is called "overconfidence." When asked to estimate an unknown but familiar quantity, respondents can be trained to provide a median estimate (the estimate about which they feel it is as likely that the true value is higher as it is that the true value is lower), as well as an interquartile range. The interquartile range is defined by the respondent's estimate of the 25th percentile (the estimate about which they feel it is 75% likely that the true value is higher and 25% likely that the true value is lower) and the respondent's estimate of the 75th percentile. For a well-calibrated respondent, it should be 50% likely that the true value would fall into the interquartile range. For example, one might be asked to give the average annual temperature in Boston, Massachusetts, USA. A respondent might provide a median estimate of 50°F, a 25th percentile estimate of 40°F, and a 75th percentile estimate of 60°F. The true average annual temperature in Boston equals 51.3°F. Were one scoring this respondent's answers, she would receive one point because her interquartile range contains the true value. A second respondent might provide a median estimate of 45°F, a 25th percentile estimate of 40°F, and a 75th percentile estimate of 50°F. Were one scoring this respondent's answers, he would receive no point because his interquartile range does not contain the true value. Note that the difference in respondents' scores derives more from the narrow width of the second respondent's interquartile range than from the distance of the median estimate from the truth. Were the second respondent's interquartile range as wide as the first respondent's (and still centered on the median estimate), then the second respondent would also have received a positive score. Setting the uncertainty range too narrowly is the hallmark of the overconfidence heuristic.

In one experiment, a cognitive scientist asked 100 students to answer ten questions like the above question about the average temperature in Boston (Alpert and Raiffa, 1982). For a well-calibrated student, one would expect the true value to lie in the interquartile range for five of the ten questions. Using the binomial distribution to set expectations, one would expect 5 or 6 of the 100 students to give answers such that 8, 9, or 10 of the true values fell into their interquartile ranges. None of the students had scores of 8, 9, or 10. One would also expect 5 or 6 of the 100 students to give answers such that 2, 1, or 0 of the true values fell into their interquartile ranges. Thirty-five of the students had scores of 2, 1, or 0. How would the distribution skew so strongly toward low scores? The skew toward low scores arises because respondents provide too narrow a range of uncertainty, so the true value lies outside the interquartile range much more often than it lies inside it. The overconfidence heuristic acts in the same way when respondents are asked to give extreme percentiles such as the 1st and 99th percentile (Alpert and Raiffa, 1982), is most pronounced when tasks are most difficult (Lichtenstein et al., 1982), has been observed to act in many different populations