10

Bayesian Statistical Methodology for Observational Health Sciences Data

Paul Gustafson

University of British Columbia, Vancouver, BC

10.1 Introduction

The data available to address important questions about harms and benefits to human health are often rather limited. Ethical concerns, logistical constraints, and resource limitations often mean that only observational studies can be conducted to address a given question about a possible determinant of health. And the data from observational studies are indeed prone to be limited in fundamental ways. Thus, while statistical methodologies play a central role across the health sciences, they become acutely critical in the observational study realm. As will be showcased here, Bayesian statistical methods can be applied to infer as much as possible about the health question at hand, while acknowledging the limitations of the available data.

As the name implies, observational studies involve measuring, but not otherwise intervening on, human subjects. It is well understood that intervention, when possible, leads to stronger evidence. Specifically, studies in which participants are randomly assigned to different levels of a potential health determinant (e.g., drug A versus drug B) lead to the firmest conclusions. Referring to a study participant's level of the potential determinant as the exposure variable, and the consequent health outcome as the disease variable, randomization yields well-supported inferences about the exposure-disease relationship. The extent to which the disease variable varies systematically across groups of participants defined by different exposure levels directly translates to a level of statistical confidence about the role of exposure in causing disease.

Unfortunately, in a wide array of settings where understanding the exposure-disease relationship is critical, randomization is impractical, unethical, or unaffordable. The next best course of action is typically an observational study, in which exposure level and health outcome are among a slew of variables measured on study participants. In such a study, however, participants are essentially self-selected into exposure groups. Thus there is far more trep-

idation in ascribing an across-group difference in health outcome as being a direct consequence of the differing exposure.

The observational studies we have in mind here fall under the rubric of "risk-factor epidemiology," addressing questions such as: Among children, does residential exposure to magnetic fields from power lines increase the risk of leukemia? Among post-menopausal women, does hormone replacement therapy (HRT) reduce the risk of osteoporosis? Among middle-aged women and men, does the "distribution" of abdominal fat, as reflected by the waist-to-hip ratio, relate to the risk of heart disease? Such questions can be subtle, and researchers strive to be exceedingly careful in how observational studies are conducted, in terms of both how the data are collected, and how statistical methods are brought to bear. However, intentions notwithstanding, messages from the research community to the public about what exposures are good and bad for health can be mixed at times, if not downright contradictory. The research community is well aware of this state of affairs; see, e.g., Taubes and Mann (1995) and Ioannidis (2005). At the end of the day, however, it can be plain old difficult to navigate from observational studies to conclusions that are sufficiently robust to convince a broad range of stakeholders and stand the scrutiny of time.

One of the biggest recent "flip-flops" in public messaging concerned HRT for post-menopausal women. Largely on the basis of evidence accrued from a series of observational studies conducted over many years, HRT became widely prescribed for relief of menopausal symptoms, and/or reduction in the risk of osteoporosis, and/or reduction in the risk of heart disease. In 2002, however, reports from a "bombshell" randomized study actually linked HRT to an elevated risk of heart disease. Prescribing practices changed abruptly, and many assigned some blame for the debacle to the limitations of observational studies.

The HRT situation is in fact quite complex. It turns out that careful consideration of the definition of exposure, with a distinction between initiation and ongoing use of HRT, goes a long way to reconciling the gap between observational study and randomized study results (Prentice et al., 2005; Hernán et al., 2008). However, some observational study challenges can be explained in a simple, pared-down context. To illustrate this, in Section 10.2 we describe a common observational study design, the case-control study. This is followed by a brief discussion of the Bayesian approach to statistical inference in Section 10.3. The Bayesian paradigm quite naturally handles situations involving multiple sources of uncertainty, as applies to many challenging observational data problems. Section 10.4 is then our "main feature." We show the Bayesian approach in action, dealing with the possibility that exposure status may not be recorded accurately for all study participants. We close in Section 10.5 with some remarks about broader applications of Bayesian methods to challenging observational data problems.

10.2 Case-Control Studies

The case-control study is a common type of observational study. A simple form arises when the exposure status and disease status of study participants are both binary. Say the exposure status of a participant is denoted as X, with $X = 0$ and $X = 1$ indicating "unexposed" and "exposed" respectively. Similarly, let $Y = 0$ and $Y = 1$ respectively indicate absence and presence of the disease in question. A case-control study proceeds by sampling n_0 members of the $Y = 0$ subpopulation to be the "controls," and n_1 members of the $Y = 1$ subpopulation to be the "cases." Note here that the study investigators set n_0 and n_1, often choosing the two sample sizes to be comparable to one another. In this way, a substantial number of the study participants can be individuals with the disease, even if the disease is actually very rare in the general population. Logistically, often this can be achieved via the use of a disease registry from which cases can be sampled. After any individual, control or case, is recruited into the study, his or her exposure status X is measured. Consequently, the fundamental summary of the study data is a "2×2 table" providing counts of study participants for each of the four combinations of exposure status X and disease status Y.

Before proceeding to an example of a 2×2 table, we must comment on a fundamental point concerning case-control studies. We seek to understand how Y depends on X. Translating to the language of conditional probabilities and conditional distributions then, we seek to learn about $\Pr(Y = y|X = x)$, where the common "bar notation" indicates conditioning, and the mathematical definition is

$$\Pr(Y = y|X = x) = \frac{\Pr(Y = y, X = x)}{\Pr(X = x)} .$$

For instance, $\Pr(Y = 1|X = 0)$ is read as the probability that $Y = 1$ given $X = 0$, and is equal to the probability that both $Y = 1$ and $X = 0$ divided by the probability that $X = 0$. On face value it seems concerning that we seek to understand the conditional distribution of Y given X, yet the data arise from the conditional distribution of X given Y, e.g., the data on controls (cases) arise according to how X values are distributed among the subset of the population for whom $Y = 0$ ($Y = 1$).

Fortunately, we can resolve this issue, particularly if we are willing to think in terms of the odds of events rather than probabilities. We would like to know to what extent the odds of contracting the disease are different for the exposed than for the unexposed. Mathematically, again using the bar notation to indicate conditioning, we can write the odds of disease given exposure status $X = x$ as

$$\text{Odds}(Y = 1|X = x) = \frac{\Pr(Y = 1|X = x)}{\Pr(Y = 0|X = x)} .$$

TABLE 10.1: Data from a fictitious case-control study with self-reported exposure status.

		Diseased ($Y = 1$)	Disease-Free ($Y = 0$)	Total
Reported Status	Exposed	150	120	270
	Unexposed	290	320	610
	Total	440	440	880

Then the Y given X odds-ratio,

$$OR_{Y|X} = \frac{\text{Odds}(Y = 1|X = 1)}{\text{Odds}(Y = 1|X = 0)},$$

summarizes how the disease rate varies between the unexposed and exposed. For instance, $OR_{Y|X} = 1.3$ would tell us that the odds of contracting the disease are 30% higher for the exposed than for the unexposed.

On first glance, since case-control data arise from the distribution of X given Y, it isn't clear that such data tell us anything about the Y given X odds-ratio. However, case-control data do directly speak to the "reverse" odds-ratio for X given Y, defined

$$OR_{X|Y} = \frac{\text{Odds}(X = 1|Y = 1)}{\text{Odds}(X = 1|Y = 0)}.$$

Then a very fortuitous bit of mathematics kicks in. In fact the two odds-ratios are identical, i.e., $OR_{Y|X} = OR_{X|Y}$, and we can unambiguously use OR to denote their common value. Consequently case-control data can indeed tell us about how disease rates, expressed on the odds scale, differ between the unexposed and the exposed.

A fictitious example of a 2×2 table for a study with $n_0 = 440$ controls and $n_1 = 440$ cases is given in Table 10.1. The labeling of the rows as giving "reported" exposure status foreshadows the problem we will dwell on presently. For now, however, we assume that the table counts really do cross-classify the study participants according to exposure status X and disease status Y. We then note a higher rate of exposure for cases than for controls. Specifically, the odds of exposure are estimated as $(150/440)/(290/440) = .517$ for cases, but only $(120/440)/(320/440) = .375$ for controls. We emphasize that while the earlier mathematical statements of odds describe the population as a whole, the data give us only estimates, since we only have information on the randomly sampled subsets on the entire control and case populations. Using the common statistical convention of the "hat-notation" to distinguish estimates from the population quantities they are estimating, $\widehat{OR} = .517/.375 = 1.38$ is then the best guess for the association between the exposure and the disease.

That is, from the available data, we have estimated the odds of acquiring the disease to be 38% higher for those population members who are exposed, relative to their unexposed compatriots.

10.3 Bayesian Analysis

In fact, we can use Bayesian statistics to formalize our inferential procedure above, and to put "error bars" around our estimated odds-ratio. The essence of the Bayesian method is to specify a prior distribution describing the state of belief about unknown quantities in advance of receiving of the data. The laws of probability then update this distribution on the basis of data received, giving the posterior distribution which summarizes the remaining uncertainty about the unknowns. In generic terms, say the unknown parameters θ are assigned a prior distribution having density function $\pi(\theta)$, and the statistical model for observable data Z given the parameters θ involves the conditional density function $\pi(z|\theta)$. Then the posterior distribution can be expressed via its density function

$$\pi(\theta|Z = z) \quad \propto \quad \pi(z|\theta)\pi(\theta),$$

with the proportionality understood to be as a function of θ. This expression is widely known as Bayes Theorem.

In the present situation, θ has two components, namely

$$\theta_0 = \Pr(X = 1|Y = 0) \quad \text{and} \quad \theta_1 = \Pr(X = 1|Y = 1),$$

the population exposure rates for controls and cases respectively. We can assign a uniform prior which doesn't favor any values for these rates over any others. The posterior distribution for θ given by Bayes' Theorem then induces the posterior distribution for the odds-ratio, since mathematically

$$OR = \frac{\theta_1/(1-\theta_1)}{\theta_0/(1-\theta_0)}\,.$$

Further, quantiles of the posterior distribution of OR can be used to report a plausible range of values in light of the data. For instance, the interval from the 5th to 95th percentile of the posterior distribution comprises a 90% credible interval for the population OR. Moreover, the theory allows a directly probabilistic interpretation of this interval: we think it nine times more likely that the true value is inside the interval rather than outside.

Based on the Table 10.1 data, and using some computational details described in the Appendix, the 90% credible interval for the population odds-ratio ranges from 1.03 to 1.90. In practical terms this is quite a lot of uncertainty: a 3% elevation in the odds of illness for the exposed is quite a different

matter than a 90% elevation. Still, at least these data speak clearly about the exposed being at elevated risk. In many studies the credible interval for the odds-ratio will straddle 1. When this happens, the data are consistent with either an elevated or diminished disease risk for the exposed.

10.4 Exposure Misclassification

With the basics of case-control studies and Bayesian inference in hand, we now admit that ascertaining the exposure status for study participants may be easier said than done. A pervasive concern in risk-factor epidemiology is that exposure status may be misrepresented for a portion of study participants. For instance, say the nature of the exposure variable in Table 10.1, perhaps along with resource constraints, limits the exposure ascertainment to self-report, e.g., each participant checks a box on a questionnaire to record his/her exposure status. This raises the possibility that a portion of participants may check the wrong box, and the consequent need to deal with this circumstance when analyzing the study data.

To delve into this concern about exposure misclassification, we introduce the technical terms of sensitivity — the probability of correct exposure classification for a truly exposed person, and specificity — the probability of correct classification for a truly unexposed person. Thus exposure misclassification is manifested if one or both of sensitivity and specificity are below 1. To keep the present example as accessible as possible, say the sensitivity can be safely assumed to be 1, but specificity may fall below 1. This could happen, for instance, if being unexposed is socially undesirable. Consider exposures such as "brushing one's teeth twice daily," or "exercising vigorously three or more times a week." With a self-report questionnaire, or even more so with a face-to-face participant interview, it is easy to imagine that all truly exposed participants report correctly. It is equally easy, however, to imagine that a portion of truly unexposed participants report incorrectly. Also to keep things simple, we assume this proportion — 1 minus the specificity — is the same for cases as it is for controls.

Pursuing this line of thinking further, if the specificity of exposure assessment is known, then the data in Table 10.1 can be analyzed accordingly. Some mathematical details are given in the Appendix. The estimated odds-ratios and 90% credible intervals based on some selected values of specificity appear in Figure 10.1. (The result corresponding to 100% specificity is identical to the "face value" result given in the previous section.) The trends seen in Figure 10.1 are in fact rather general. As we assume lower specificity, we assume there is more random variation in our Table 10.1 data. Such variation will weaken, or attenuate, the relationship between reported exposure and disease. Thus as we assume more randomness is present, we should boost, or

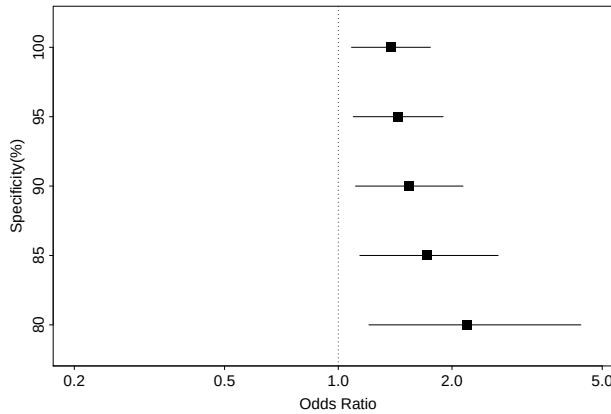

FIGURE 10.1: Inference for the exposure-disease odds-ratio based on Table 10.1 data, for selected values of specificity. The posterior estimate of the odds-ratio and the 90% equal-tailed credible interval are displayed. Technically, the posterior estimate is the exponentiated posterior mean of the log odds-ratio. Note the use of a logarithmic axis.

de-attenuate, the estimated odds-ratio for the actual exposure and disease. The upward march in the estimated odds-ratio as the assumed specificity is reduced is then intuitively anticipated. By the same token, admitting that the data are more afflicted by random error is tantamount to admitting that they contain less useful information. Therefore, we are not surprised to see the credible interval widen as the assumed specificity decreases. Finally, it is both interesting and typical that in some sense the two trends seen cancel each other out. As the assumed specificity decreases, the increase in estimated odds-ratio and the widening of the credible interval combine such that the lower endpoint of the credible interval is actually quite stable.

Reflecting upon Figure 10.1, we have five different statistical inferences that adjust for exposure misclassification, corresponding to five different assumptions about the magnitude of the misclassification. If the only overarching judgment we are comfortable with is that the specificity is 80% or more, then perhaps the ensemble of inferences is the take-away item. Such a "what-if" analysis is often referred to as a sensitivity analysis (as in sensitivity of the inference to the assumption, not the technical sense of sensitivity as a probability of correct classification), or a stress test. Of course the results of a sensitivity analysis are disquieting when, as at present, the inference changes considerably with the assumption.

Intuitively, at least, our Table 10.1 data do not appear to tell us anything about the extent of exposure misclassification. Since we don't observe the true exposure status for any individuals, we cannot infer how commonly the

TABLE 10.2: Data from a second fictitious case-control study with self-reported exposure status.

		Diseased ($Y = 1$)	Disease-free ($Y = 0$)	Total
Reported Status	Exposed	188	168	356
	Unexposed	692	712	1404
	Total	880	880	1760

true status and self-reported status are discordant. We don't have to rely on intuition here, however. Just as Bayesian inference yields the posterior distribution of the target parameter given the data and an assumed value of specificity, it can also give us the posterior distribution of the specificity itself. As always with Bayesian inference, such calculation of posterior probabilities is predicated on the assignment of prior probabilities. Say, for instance, we assign equal prior probabilities of .20 to each of the five specificity values (80%, 85%, ..., 100%) considered in Figure 10.1. The Table 10.1 data then induce corresponding posterior probabilities of .248, .220, .196, .176, .159 (again, some of the calculation details appear in the Appendix). Thus the data are very slightly more suggestive of a smaller value of specificity. This "tilt" is so small though, that we are really just left with the what-if message from Figure 10.1. This is a fairly common situation. We can posit different assumptions about the extent of a problem with observational data, while the data themselves have little to commend one assumption over the others.

Our plot thickens upon considering a second fictitious dataset, given in Table 10.2. The what-if analysis for these data appears in Figure 10.2. Again, as we lower the assumed specificity, the best-guess odds-ratio is boosted, and the associated credible interval widens. This time, however, the credible interval consistently crosses 1. Thus we are not completely convinced that exposure is associated with an elevation in disease risk, regardless of how much exposure classification error is assumed to be manifested in the data. Also, note that the change in inference upon reducing the assumed specificity from 85% to 80% is particularly marked.

Again we can start with equal prior probabilities for the five specificity values and then update these based on the data. This results in posterior probabilities of .068, .273, .243, .218, .197, for specificity ranging from 80% up to 100% by steps of 5%. A specificity of 80% is thus far less supported by the data than the four higher values. In one sense this is good news. Of the five analyses, the one that stands furthest apart from the others is also the one given by far the least credence by the data. Against this, however, the story has become more complex. For some datasets, such as that in Table 10.1, a sensitivity analysis is more or less just that: inferences under some different scenarios are reported, and the a priori uncertainty about which scenarios

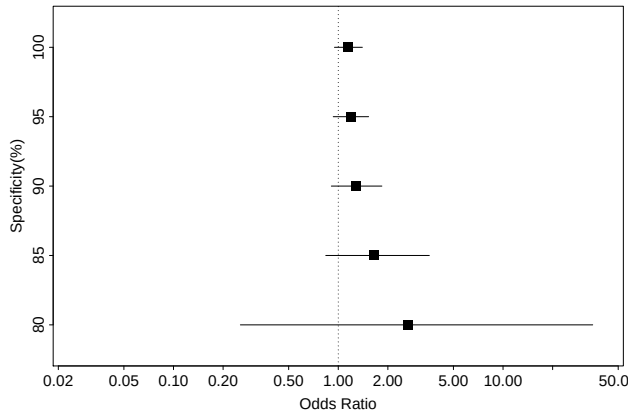

FIGURE 10.2: Inference for the exposure-disease odds-ratio based on Table 10.2 data, for selected values of specificity. The posterior estimate of the odds-ratio and the 90% equal-tailed credible interval are displayed. Technically, the posterior estimate is the exponentiated posterior mean of the log odds-ratio. Note the use of a logarithmic axis.

are most plausible remains largely unchanged upon receipt of the data. Other datasets, however, have considerably more to say, e.g., one of the five inferences in Figure 10.2 is considerably less plausible than the others. Also, note that for a new dataset coming onto the scene, we won't know which situation applies until we carry out the posterior analysis.

As a final thought about these examples, once we have the posterior distribution of the odds-ratio for each postulated value of specificity, as well as the posterior distribution of the specificity, the laws of probability dictate how the five what-if posterior analyses can be synthesized into an "overall" posterior analysis. Applied to the first dataset, this synthesis produces a best estimate of 1.67 for the odds-ratio, with a 90% credible interval running from 1.11 to 3.02. With reference to Figure 10.1, this looks about right for bringing the five inferences together. For the second dataset, we synthesize a best guess of 1.38 for the odds-ratio, with a 90% credible interval from .89 to 3.07. With reference to Figure 10.2, we see that the lower posterior probability attached to 80% specificity results in a synthesis relying less on this scenario than the others.

10.5 Going Further with Bayesian Analysis

We have seen through a simple, if somewhat contrived example of using Bayesian analysis to acknowledge a common limitation of observational study data. In the context of a simple exposure variable prone to a simple form of misclassification, Bayesian inference was used to estimate the exposure-disease association while acknowledging uncertainty about the extent of misclassification. We emphasize that in this sort of problem we are indeed mitigating, rather than repairing, the damage caused by the misclassification. No amount of statistical cleverness can recover the actual exposure status for each study participant and thereby reclaim the inference we would ideally have drawn. At best we can draw inferences with sufficiently wide credible intervals to acknowledge the damage done by misclassification.

In fact, issues of poor measurement in observational studies can arise in a much more complex manner, in the face of much more complex data structures. Moreover, measurement error is often mentioned as but one of three major challenges that make it hard to obtain broadly convincing evidence from an observational study. Measurement error, along with uncontrolled confounding and selection bias, have been referred to as a "holy trinity" of threats to validity for observational studies in risk-factor epidemiology (Greenland, 2009). Roughly, uncontrolled confounding refers to the complete absence of data on a crucial variable that is associated with exposure status and also associated with disease status. For instance, it may only make sense to look at the relationship between abdominal fat distribution and heart disease within subpopulations defined by a genotype, but if this genotype is not recorded in the available data, then we have a severe challenge on our hands. Selection bias, on the other hand, is somewhat self-explanatory. If the characteristics of study participants somehow deviate systematically from the characteristics of the population as a whole, then we need statistical analysis which acknowledges this lack of representation.

Statisticians are kept busy with devising methodology to deal with this trinity of threats, and others, in the observational study realm. The Bayesian approach provides a general framework for such methodologies. As in Section 10.4, but with more complex data structures, one can estimate a target quantity from less than ideal data, and report an honest amount of uncertainty surrounding the estimate. Arguably, the Bayesian paradigm is particularly good for this uncertainty management. For instance, the credible intervals reported at the very end of Section 10.4 seamlessly integrate three types of uncertainty. The first type is the usual statistical uncertainty arising from having measurements on only the study participants (a sample), not the whole population. The second type is the additional uncertainty arising from having less than ideal data on a per participant basis, e.g., having self-reported exposure status rather than actual exposure status. The third type is the yet

further uncertainty arising because the extent to which the data are less than ideal is itself unknown. For instance, a priori we might plausibly think that the specificity of the exposure classification could be as low as 80% or as high as 100%, and, as we have seen, this prior belief may not be changed very much upon observing the data.

Also with regard to the trinity of threats to validity for observational studies, it is important to note that two of these threats, or even all three, can easily manifest themselves in the same study. Consequently, some of the most impressive work is that developing Bayesian (or nearly Bayesian) schemes is to tackle multiple failings of the data simultaneously. Great examples of this are the methodological developments of Greenland (2003, 2005), which are applied to observational studies concerning magnetic fields from power lines and childhood leukemia.

This volume celebrates the International Year of Statistics from a Canadian perspective. Happily, in the realm of Bayesian methods for observational study data, there is considerable Canadian activity to report. Researchers in Montréal and Vancouver are particularly active in this area. On unobserved confounding, see, e.g., McCandless et al. (2007, 2008), Gustafson et al. (2010), and McCandless (2012). Concerning measurement error, recent work includes Ladouceur et al. (2007), Hossain and Gustafson (2009), Liu et al. (2009), Dendukuri et al. (2010), Chu et al. (2010), Lu et al. (2010), and Espino-Hernandez et al. (2011), and Wang et al. (2012).

A considerable portion of the current excitement and challenge and societal impact in statistical science lies toward the "Big Data" frontier. While it may be a quieter cousin, the "Difficult Data" frontier also generates excitement and challenge and societal impact. Some of the most important decisions to be taken, such as those relating to what exposures help or harm our health, are informed by some of the most difficult data. As part of the modern statistical firmament, Bayesian inference applied to observational study data exemplifies the principled extraction of evidence from difficult data.

Appendix

Let C_0 and C_1 be the counts of apparent exposures among the n_0 controls and n_1 cases respectively. The analyses presented in Sections 10.3 and 10.4 arise from modeling these counts as binomially distributed. That is, C_i counts the number of "successes" among n_i independent trials, each of which has success probability θ_i^*. In formal notation, $C_i \sim \text{Bin}(n_i, \theta_i^*)$ independently for controls $(i = 0)$ and cases $(i = 1)$.

For given sensitivity SN and specificity SP of exposure classification, one has $\theta_i^* = \theta_i SN + (1 - \theta_i)(1 - SP)$, where $\theta_i = \Pr(X = 1|Y = i)$ is the true proportion exposed. Assigning the prior distribution $\theta_i \sim \mathcal{U}(0, 1)$, indepen-

dently for $i = 0, 1$ induces the prior $\theta_i^* \sim \mathcal{U}(1 - SP, SN)$. Consequently, the posterior distribution of θ_i^* is the $\mathcal{B}(c_i + 1, n_i - c_i + 1)$ distribution truncated to the interval $(1 - SP, SN)$, again independently for $i = 0$ or 1. A Monte Carlo sample can then be drawn directly from the posterior distribution of (θ_0^*, θ_1^*) and transformed to realizations from the (θ_0, θ_1) posterior.

To compare the posterior weight given to different values of (SN, SP), one must compute the marginal probability $\Pr(C_0 = c_0) \times \Pr(C_1 = c_1)$ for each pair of values considered. The desired quantity can be computed via a routine for the Beta distribution, given that $\Pr(C_i = c_i)$ can be written as

$$\frac{1}{SN + SP - 1} \left(\begin{array}{c} n_i \\ c_i \end{array} \right) \int_{1-SP}^{SN} \theta_i^{*(c_i+1)-1} (1 - \theta_i^*)^{(n_i - c_i + 1)-1} \, \mathrm{d}\theta_i^*.$$

About the Author

Paul Gustafson is a professor of statistics at the University of British Columbia. He received his undergraduate training at UBC and a PhD from Carnegie Mellon University, Pittsburgh, PA. His research interests include Bayesian inference, causal inference, epidemiology, measurement error, and the analysis of observational data. He has served as editor of *The Canadian Journal of Statistics* (2007–09) and was the 2008 winner of the CRM–SSC Award. He is a fellow of the American Statistical Association.

Bibliography

Chu, R., Gustafson, P., and Le, N. (2010). Bayesian adjustment for exposure misclassification in case-control studies. *Statistics in Medicine*, 29:994–1003.

Dendukuri, N., Bélisle, P., and Joseph, L. (2010). Bayesian sample size for diagnostic test studies in the absence of a gold standard: Comparing identifiable with non-identifiable models. *Statistics in Medicine*, 29:2688–2697.

Espino-Hernandez, G., Gustafson, P., and Burstyn, I. (2011). Bayesian adjustment for measurement error in continuous exposures in an individually matched case-control study. *BMC Medical Research Methodology*, 11:67.

Greenland, S. (2003). The impact of prior distributions for uncontrolled confounding and response bias: A case study of the relation of wire codes and magnetic fields to childhood leukemia. *Journal of the American Statistical Association*, 98:47–55.

Greenland, S. (2005). Multiple-bias modelling for analysis of observational data. *Journal of the Royal Statistical Society, Series A*, 168:267–306.

Greenland, S. (2009). Accounting for uncertainty about investigator bias: Disclosure is informative how could disclosure of interests work better in medicine, epidemiology and public health? *Journal of Epidemiology and Community Health*, 63:593–598.

Gustafson, P., McCandless, L. C., Levy, A. R., and Richardson, S. (2010). Simplified Bayesian sensitivity analysis for mismeasured and unobserved confounders. *Biometrics*, 66:1129–1137.

Hernán, M. A., Alonso, A., Logan, R., Grodstein, F., Michels, K. B., Willett, W. C., Manson, J. E., and Robins, J. M. (2008). Observational studies analyzed like randomized experiments: An application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology*, 19:766–779.

Hossain, S. and Gustafson, P. (2009). Bayesian adjustment for covariate measurement errors: A flexible parametric approach. *Statistics in Medicine*, 28:1580–1600.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2:e124.

Ladouceur, M., Rahme, E., Pineau, C. A., and Joseph, L. (2007). Robustness of prevalence estimates derived from misclassified data from administrative databases. *Biometrics*, 63:272–279.

Liu, J., Gustafson, P., Cherry, N., and Burstyn, I. (2009). Bayesian analysis of a matched case-control study with expert prior information on both the misclassification of exposure and the exposure-disease association. *Statistics in Medicine*, 28:3411–3423.

Lu, Y., Dendukuri, N., Schiller, I., and Joseph, L. (2010). A Bayesian approach to simultaneously adjusting for verification and reference standard bias in diagnostic test studies. *Statistics in Medicine*, 29:2532–2543.

McCandless, L. C. (2012). Meta-analysis of observational studies with unmeasured confounders. *International Journal of Biostatistics*, 8(2):art. 5.

McCandless, L. C., Gustafson, P., and Levy, A. R. (2007). Bayesian sensitivity analysis for unmeasured confounding in observational studies. *Statistics in Medicine*, 26:2331–2347.

McCandless, L. C., Gustafson, P., and Levy, A. R. (2008). A sensitivity analysis using information about measured confounders yielded improved assessments of uncertainty from unmeasured confounding. *Journal of Clinical Epidemiology*, 61:247–255.

Prentice, R. L., Langer, R., Stefanick, M. L., Howard, B. V., Pettinger, M., Anderson, G., Barad, D., Curb, J. D., Kotchen, J., Kuller, L., et al. (2005). Combined postmenopausal hormone therapy and cardiovascular disease: Toward resolving the discrepancy between observational studies and the women's health initiative clinical trial. *American Journal of Epidemiology*, 162:404–414.

Taubes, G. and Mann, C. C. (1995). Epidemiology faces its limits. *Science*, 269:164–169.

Wang, D., Shen, T., and Gustafson, P. (2012). Partial identification arising from non-differential exposure misclassification: How informative are data on the unlikely, maybe, and likely exposed? *International Journal of Biostatistics*, 8(1):Article 31.