# Editorial

<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>185</td>
<td>Virtual Communities</td>
<td>Allen J. Wilcox</td>
</tr>
</tbody>
</table>

# Commentary

<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>186</td>
<td>Dissecting Effects of Complex Mixtures: Who's Afraid of Informative Priors?</td>
<td>Duncan C. Thomas; John S. Witte; Sander Greenland</td>
</tr>
<tr>
<td>191</td>
<td>Right or Wrong?: On the Difficult Relationship Between Epidemiologists and Handedness.</td>
<td>Olga Basso</td>
</tr>
<tr>
<td>194</td>
<td>Statins and Cancer Risk: What Do We Know and Where Do We Go From Here?</td>
<td>Patricia G. Moorman; Robert J. Hamilton</td>
</tr>
<tr>
<td>197</td>
<td>Maternal Smoking and Undescended Testes: Reaching a Tipping Point.</td>
<td>Martha Werler</td>
</tr>
</tbody>
</table>

# Original Article

<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>199</td>
<td>Bayesian Methods for Highly Correlated Exposure Data.</td>
<td>Richard F. MacLehose; David B. Dunson; Amy H. Herring; Jane A. Hoppin</td>
</tr>
<tr>
<td>208</td>
<td>Innate Handedness and Disease-Specific Mortality in Women.</td>
<td>Made K. Ramadhan; Sjoerd G. Elias; Paulus A. H. van Noord; Diederick E. Grobbee; Petra H. M. Peeters; Cuno S. P. M. Uiterwaal</td>
</tr>
<tr>
<td>213</td>
<td>Statin Use and the Risk of 10 Cancers.</td>
<td>Patricia F. Coogan; Lynn Rosenberg; Brian L. Strom</td>
</tr>
<tr>
<td>220</td>
<td>Cryptorchidism According to Maternal Gestational Smoking.</td>
<td>Morten Søndergaard Jensen; Gunnar Toft; Ane Marie Thulstrup; Jens Peter Bonde; Jørn Olsen</td>
</tr>
<tr>
<td>226</td>
<td>Maternal Smoking and Environmental Tobacco Smoke Exposure and the Risk of Orofacial Clefts.</td>
<td>Margaret A. Honein; Sonja A. Rasmussen; Jennita Reehuis; Paul A. Romitti; Edward J. Lammer; Lixian Sun; Adolfo Correa</td>
</tr>
<tr>
<td>234</td>
<td>Prepregnancy Body Mass Index and the Occurrence of Severe Hypertensive Disorders of Pregnancy.</td>
<td>Lisa M. Bodnar; Janet M. Catov; Mark A. Klebanoff; Roberta B. Ness; James M. Roberts</td>
</tr>
</tbody>
</table>
Variation in Incidence of Neurodevelopmental Disorders With Season of Birth.
Hjörðís Ósk Atladóttir; Erik T. Parner; Diana Schendel; Søren Dalsgaard; Per Hove Thomsen; Poul Thorsen

Guided Multiple Imputation of Missing Data: Using a Subsample to Strengthen the Missing-at-Random Assumption.
Gary Fraser; Ru Yan

Alternative Methods of Estimating an Incubation Distribution: Examples From Severe Acute Respiratory Syndrome.
Benjamin J. Cowling; Matthew P. Muller; Irene O. L. Wong; Lai-Ming Ho; Marie Louie; Allison McGeer; Gabriel M. Leung

Virtual Epidemiology

Modeling Infectious Diseases Dissemination Through Online Role-Playing Games.
Ran D. Balicer

Self Study

The Fate of Epidemiologic Manuscripts: A Study of Papers Submitted to Epidemiology.
Susan A. Hall; Allen J. Wilcox

Brief Report

Scott Davis; Dana K. Mirick

Ambient Manganese Exposure is Negatively Associated With Human Sperm Motility and Concentration.
Julia J. Wirth; Mary G. Rossano; Douglas C. Daly; Nigel Paneth; Elizabeth Puscheck; Rachel C. Potter; Michael P. Diamond

Is Height Associated With Cardiovascular Risk in Chinese Adults?
C Mary Schooling; G Neil Thomas; Gabriel M. Leung; Sai Yin Ho; Edward D. Janus; Tai Hing Lam

Vignette

Austin Flint, Clinician Turned Epidemiologist.
Warren Winkelstein Jr

Letters to the Editor

Trends in Obesity in Pregnancy.
Joel G. Ray; Rosane Nisenbaum; Gita Singh; Chris Meier; Andrea Guerin; Philip R. Wyatt; Marian J. Vermeulen

Coffee and Myocardial Infarction.
Andreas J. Flammer; Isabella Sudano; Roberto Corti

Coffee and Myocardial Infarction.
Viktor uli
282 **Coffee and Myocardial Infarction.**
Ana Baylin; Sonia Hernandez-Diaz; Edmond K. Kabagambe; Xinia Siles; Hannia Campos

283 **Calculating Disease Burden.**
Peter Morfeld

283 **Calculating Disease Burden.**
Ben Armstrong; Kyle Steenland

284 **Breast Cancer Trends.**
Noel S. Weiss

284 **Breast Cancer Trends.**
Robert E. Tarone
Virtual Communities

Allen J. Wilcox

Scientific journals, like newspapers, help create a sense of community. The community that reads *Epidemiology* is not joined by geography or family relationships, but it is just as real—a community that nourishes friendships, entertains feuds, celebrates reunions, and shares common values across distance and culture.

One place where this virtual community is expressed most concretely is in our Editorial Board. The members of the Board take on a special responsibility to the journal. They provide extra duty as frequent reviewers. They make the final selection for the annual Rothman *Epidemiology* Prize (with a new winner to be announced in our next issue). Perhaps most importantly, they act as the editors’ eyes and ears in the wider world, providing us with suggestions for the journal and offering feedback.

When the present Editors and I assumed stewardship of the journal 6 years ago, we instituted 5-year terms for our Board members. This year we say thanks to a particularly distinguished lot as they finish their terms: Pierre Buekens, Sander Greenland, Patricia Hartge, Irva Hertz-Picciotto, and Charlie Poole. Every one has provided indispensable service to the journal during these past years. Sander deserves special recognition for his long run—he was on the inaugural Editorial Board established by Ken Rothman at the journal’s founding in 1990. His record of service with *Epidemiology* is unlikely to be surpassed.

We are pleased to announce 5 new members to our Board: Sven Cnattingius of Karolinska Institutet, Stephen Cole of Johns Hopkins University, Sandro Galea and Hal Morgenstern, both at the University of Michigan, and Miquel Porta of the Universitat Autònoma de Barcelona. To these distinguished colleagues we say welcome, and thanks for helping the journal extend its connections to new neighborhoods.
**Dissecting Effects of Complex Mixtures**

**Who’s Afraid of Informative Priors?**

**Duncan C. Thomas,* John S. Witte,† and Sander Greenland‡**

Abstract: Epidemiologic studies commonly investigate multiple correlated exposures, which are difficult to analyze appropriately. Hierarchical modeling provides a promising approach for analyzing such data by adding a higher-level structure or prior model for the exposure effects. This prior model can incorporate additional information on similarities among the correlated exposures and can be parametric, semiparametric, or nonparametric. We discuss the implications of applying these models and argue for their expanded use in epidemiology. While a prior model adds assumptions to the conventional (first-stage) model, all statistical methods (including conventional methods) make strong intrinsic assumptions about the processes that generated the data. One should thus balance prior modeling assumptions against assumptions of validity, and use sensitivity analyses to understand their implications. In doing so—and by directly incorporating into our analyses information from other studies or allied fields—we can improve our ability to distinguish true causes of disease from noise and bias.

(Epidemiology 2007;18: 186–190)

Multiple regression techniques and the broader class of generalized linear models have a long history in statistics and are widely used in epidemiology. Recent years have seen important frequentist and Bayesian advances in hierarchical and multilevel generalized linear models (ie, generalized linear mixed models). These approaches provide a unified framework to tackle many problems that are not well addressed by standard statistical models, particularly issues involving various levels of clustering in one’s data. Their traditional use in social sciences is to address problems involving correlated outcomes, as arise in cluster sampling designs.

Of perhaps greater interest to epidemiologists is the use of hierarchical models to study highly correlated exposures, as investigated by MacLehose et al in this issue. Here, one can improve exposure effect estimates by modeling them in a higher-level model. MacLehose and colleagues assume that exposures with similar effect sizes arise from the same prior distribution. One can also more directly incorporate external information about the exposures into a higher-level model. For example, one could use information on the sources of air pollution components or their interactions with metabolizing enzymes as a way of dissecting which constituents are etiologically relevant or which sources are responsible. Other examples of external information include exposures’ chemical structure or composition, nutrient content, or genomic properties.

Another issue addressed by hierarchical models has arisen with the “-omics” revolution: the “large p, small n” problem, namely many more variables (p) than observations (n). Standard multiple regression techniques break down here because some combination of variables can always be found that will perfectly predict any other variable (ie, the design matrix is singular). Models are also becoming more complex, as investigators seek to understand biologic pathways leading from multiple exposures to disease through intermediates that are influenced by various genes. Hierarchical models can overcome these problems by modeling many fewer variables that retain characteristics of the original p exposures.

Moreover, techniques like hierarchical and Bayesian model averaging can be used to reflect the uncertainty in model form. This is especially important because a single model is seldom fitted in the course of data analysis; yet all too often, the estimates from a single “best” model are presented along with confidence limits that do not reflect the uncertainty about the form of the model. Hierarchical model averaging, in particular, allows inclusion of all candidate variables in the model and thus eliminates the need for variable selection and its attendant artifacts.

Observational studies are also subject to many different types of potential bias, which are often considered in a qualitative manner in the discussion sections of papers or treated in some form of sensitivity analysis. But as we will discuss below, Bayesian methods are available for formally modeling the different sources of bias.

**HIERARCHICAL MODELING OF MULTIPLE EXPOSURES**

MacLehose and colleagues describe broad classes of higher-level (prior) distributions for incorporating similarities among multiple exposures in their analysis. Two of the prior distributions are fully parametric and 2 are semiparametric,
and all can include a nonzero probability mass at the null hypothesis of no exposure effect.

Parametric models have the advantage that covariates are readily incorporated in the prior distribution. These covariates define “exchangeability classes,” a priori groupings of variables that are assumed to have equivalent effects on disease (i.e., that derive from the same prior distribution). The specification of such groupings does not, however, require the investigator to specify exactly how such groups might differ or even to declare that they really do differ. They merely specify groups that potentially differ and the data will indicate whether this grouping has any predictive value.

Although the presentation by MacLehose et al of prior covariates as defining a set of distinct categories is simple, the general hierarchical modeling framework does not require discrete groupings, and indeed one can incorporate a whole vector of prior covariates on continuous or categorical scales. Thus, an investigator does not need to define explicit prior probabilities, but rather a set of prior covariates that might define distinguishable subgroups, and the general form of the prior probability distribution (along with its dependence on the covariates or “hyperparameters”). For this approach to provide notable statistical benefits, the number of effects under study (first-stage parameters) must be substantially larger than the number of prior (second-stage) parameters to be estimated.

MacLehose and colleagues do not focus on use of prior knowledge, although they mention using broad classes of pesticides as a grouping variable. Instead, they consider the case in which one suspects some meaningful similarities among the variables but has no idea what they are. They employ a semiparametric Dirichlet process prior, which allows each effect under study to come from a discrete but unknown distribution. Different effects may come from the same distribution, but the grouping of effects according to their shared distribution is also unknown. Ultimately, effects are grouped by similarity in estimated size rather than by external properties (such as chemical structure), and each effect estimate is shrunk toward effects of similar initial size. Such an analysis also allows other kinds of inferences, such as estimation of the posterior probability of that a given effect is null, and the posterior probability that 2 given variables belong to the same cluster.

The flexibility of semiparametric methods is seductive, but has the drawback of allowing (and often producing) possibilities that are simply not credible in most applications, such as multiple modes and groupings that contain physically dissimilar exposures. The semi-parametric model thus represents the opposite extreme of rigid two-parameter priors such as the normal distribution, which offer essentially no flexibility apart from location and scale. Such priors are generally chosen for convenience (ease of representation) or ignorance (we want a continuous distribution but have only a vague idea what shape to use). If an investigator is seriously worried about such assumptions, however, they can perform a sensitivity analysis. Alternatively, a more flexible parametric class of distributions can be used that can approximate reasonable possibilities such as skewness and heavy tails, but conforms to desired constraints such as smoothness, unimodality, and strict monotonicity. An example is 4-parameter generalized conjugate (log-F) family of distributions.

To summarize our point, one should appreciate that the Dirichlet process prior model in the paper by MacLehose et al is not using any external information to derive the grouping, in contrast to the use of prior covariates in a hierarchical model. Instead, exposures are clustered together based on their observed effect sizes, so the resulting clusters need not have anything else in common. This can be helpful when there are outliers that really should not be grouped with the bulk of the data. Nonetheless, it can also accentuate spurious clustering that has no biologic basis but instead merely reflects chance similarities. Fortunately, in their application to the pesticide data, this does not seem to have happened, as the observations are generally shrunk towards a highly concentrated distribution quite near the null.

**SHRINKAGE, VARIABLE SELECTION, OR MODEL AVERAGING?**

The traditional approach to dealing with multiple covariates is some form of variable selection, such as stepwise regression. As is well known, such techniques can lead to biased point and interval estimates and generally do a poor job of choosing the right variables, particularly when the variables are highly correlated, although they may produce reasonable prediction models. Even for the latter purpose, however, stepwise regression generally yields less accurate prediction equations than putting all variables into the equation and using hierarchical modeling techniques to stabilize their coefficients.

One objection to the hierarchical approach is that it does not yield a parsimonious model, if only because all variables are included. This objection has led to the development of a variety of techniques that include some form of variable selection, often in a Bayesian framework. Combined with frequentist or Bayesian model averaging, one can then make statements about both the relative contributions of the different variables and the coefficients of the variables, allowing for uncertainty about which other variables to adjust for. As with hierarchical models, these averaged models generally yield better predictions than conventional variable-selection methods; however, they are admittedly more complex, as again all variables must be considered. Although these techniques can be applied without use of external data, they can easily incorporate prior information, either in models for the regression coefficients in each model, or in the model for the selection of variables, or in both. For example, Conti et al described a hierarchical framework for modeling complex metabolic pathways. They used a logistic model for the probability of a nonzero coefficient for each gene effect, environmental factor effect, and product among them (“interaction”) and a linear model for the magnitude of the coefficient given it was nonzero. This approach also provides a natural way of exploiting any hierarchical structure of the covariates themselves, as in the analysis of SNPs within linkage disequilibrium blocks within genes within pathways.
WHY HAVE HIERARCHICAL MODELS SEEN LIMITED USE IN EPIDEMIOLOGY?

Considering the potential applications and value of hierarchical models, it is surprising that they are not used more often. One explanation is their computational difficulty, but this argument is obsolete in light of the growing availability of software and approaches such as data augmentation priors for hierarchical modeling. With the latter, prior information is expressed as “prior data” and added to the observed data, allowing Bayesian analysis with conventional software.

This leads to another reason voiced for not using hierarchical models: they require thoughtful application and thus can be difficult to specify. A key to their proper use is understanding that problems are solved by the introduction of external information concerning properties of and relations among the variables. This information is most needed—and potentially most helpful and most hazardous—when the data under analysis have little or no information about these relations. The quality of the external information then becomes as central to the analysis as the quality of the data themselves.

There may also be some stigma about Bayesian methods reflecting a misguided belief that inferences were entirely dependent upon the user’s subjective choices of priors. This probably reflects a basic misunderstanding about how hierarchical priors are specified. As described earlier, these priors only define potentially distinguishable exchangeability classes, not the magnitude of differences among the classes. The False Positive Report Probability was presented as a means of combining an investigator’s prior belief about the probability that the null hypothesis was false with observed data to obtain a posterior probability by straightforward application of Bayes formula. Of course, an investigator who was uncomfortable with the idea of specifying a single number for that prior probability could always redo the calculation over a range of priors to obtain a range of posterior probabilities, but reducing the problem to a simple binary decision to accept or reject the null hypothesis is arguably unsatisfactory. More appropriate would be to consider the entire range of possible sizes for (say) a relative risk, give this range a prior distribution informed by external information, and see how the data modifies that distribution.

SHOULDN’T WE LET DATA “SPEAK FOR THEMSELVES”?

These issues highlight the need for reasonable assumptions and high-quality external information in hierarchical modeling. The latter may sound daunting, but is in fact no different than the quality of external information demanded for more common tasks, such as proper selection of variables for confounder control (e.g., do not control variables affected by exposure). More generally, the harsh reality is that data by themselves say nothing at all; every inference, even from nonparametric methods, depends crucially on sound judgment and external information. Models are always necessary to provide a context for interpreting the data. This is even true for randomized trials, where the results are driven by assumptions about noncompliance and drop out, and that there was no ad-hoc data manipulation—assumptions that are not always true (as the recent Vioxx controversy reminds us). Every analysis that fails to explicitly model a potential source of bias assumes that the bias is absent. In most observational data analyses, few biases are modeled. It is thus no surprise that underestimation of biases is a recurring problem, as refutation of observational results by controlled trials reminds us. Some statisticians talk of analyses being “data-driven” almost virtuously when in fact it may be a bad thing in some contexts, because it may be the data and not the models or priors that are wrong.

The challenge is striking a balance between validity (zero-bias) assumptions and the external contextual information that we inject as modeling assumptions or priors. Because the optimal balance is never known, we need to do sensitivity analyses in which the relative strength of the 2 are varied. To the extent that more sophisticated methods like those of MacLehose et al allow us to explore further options, they are a valuable addition to our toolkit. The danger is only in presuming inherent superiority because of adjectives like “nonparametric” and “semiparametric.” These descriptors are often taken as virtues, but should really be a red flag, because they indicate that the data will be taken as unquestionably superior to any model along certain dimensions. It is not hard to see such methods break down in simple examples in which the data are neither pure nor ample. A nonparametric curve going wild at the ends is a nice way to show that even with perfect random sampling we always want to rely somewhat on prior information, such as monotonicity and smoothness.

At the other extreme, we have methods for bias modeling that do not take the data as completely trustworthy, and so are heavily model-driven. Because these methods admit uncertainties along dimensions that are not even identified by the data, they cannot be nonparametric; they rely entirely on investigator inputs about validity problems. These Bayesian approaches view the data as nothing but an observed margin of a much higher-dimensional distribution, and one that can be practically connected to the desired effects only by using parametric models and priors. As discusses exist in the realm of nonidentified modeling, meaning they do not make assumptions strong enough to produce point estimates unless one introduces explicit priors (as opposed to the zero-bias priors implicit in conventional methods). This realm is arguably far more appropriate for observational study analysis than assuming biases away.

From the nonidentified modeling perspective, nonparametric and semiparametric methods are modeling only a marginal distribution that may have tenuous connection to our real targets. The latter targets may be buried deep in the...
full joint distribution of biases and effects. A fair comparison of nonparametric and semiparametric methods with parametric methods would evaluate them both, allowing for uncertainty due to uncontrolled data problems. When this is done, both approaches will be seen to understate the uncertainty warranted by our limited background understanding. The differences in mean squared error of nonparametric and semiparametric methods relative to parametric methods will no longer be so dramatic, and may easily favor the parametric approach, depending critically on the prior used to evaluate the nonidentified bias components.

This comparison also bears on the issue of the practicality of effort needed to set up and interpret each method. We cannot do everything when analyzing a study. Where should our limited analytic resources be focused if we want to most improve our overall performance? Many if not most epidemiologists think they can deal with biases informally and intuitively, saving them the effort of formal bias modeling. Meanwhile, massive computing power has enabled statisticians to implement dazzling and sophisticated techniques for dealing with random error. To the extent that biases are due to use of parametric models rather than fundamental study problems, these methods can help, but no amount of sophistication in this regard can compensate for study problems. The growing list of topics in which informal validity judgments have proven wrong (eg, beta-carotene and cancer prophylaxis, hormone replacement therapy and cardioprotection) indicates that we need to do better in evaluating the strength (or rather, weakness) of the epidemiologic evidence when health policy and medical practice is at stake. Reanalysis of the epidemiologic data from past controversies might shed more light on the merits of various approaches than would analytic or simulation studies. It would also help inform us about how much analytic sophistication—as opposed to more sophisticated data collection, as in “validation studies”—could have saved us from past mistakes.

CONCLUSION

In the end, we use informative priors, whether we admit it to ourselves and others or not. This is true even for so-called “objective” methods such as conventional frequentist methods and objective-Bayesian methods. How informative our conclusions appear will depend entirely on how informative our analysis assumptions are, even if we have billions of subjects. It is therefore of utmost importance that we recognize and admit those assumptions and make the best effort to found them on sound scientific information. Nor should we miss any opportunity to exploit what information may be available from other studies or allied fields, for this information will help us better discern signal from noise and bias in our data.

ABOUT THE AUTHOR

DUNCAN C. THOMAS is Verna Richter Professor in the Department of Preventive Medicine at the University of Southern California. JOHN S. WITTE is Professor of Epidemiology and Biostatistics and the Associate Director of the Institute for Human Genetics at UC San Francisco. SANDER GREENLAND is Professor of Epidemiology and Statistics at UC Los Angeles. Drs. Thomas, Witte, and Greenland have longstanding research interests in epidemiologic methods, including the development and application of hierarchical models. Over the years they have enjoyed numerous collaborations.

REFERENCES

42. Good IJ. *Good Thinking*. Minneapolis: University of Minnesota Press; 1983.
Right or Wrong?

On the Difficult Relationship Between Epidemiologists and Handedness

Olga Basso

Abstract: The declining prevalence of left-handers with age has resulted in the hypothesis that sinistrality, being the result of a developmental insult, may be associated with a reduced life span. While it is plausible that some individuals become left-handed as a consequence of neurologic impairment, the literature on handedness itself appears to suffer from a number of problems. These include the ease with which information on handedness can be collected in the absence of prior hypotheses, the failure to address heterogeneity among left-handers, and the selective publication of positive results. Even if individual contributions, including one published in this issue of Epidemiology, are of reasonable quality, all the above problems conspire to lower the credibility of this area of research.

(Epidemiology 2007;18: 191–193)

Dextrals, also known as right-handers, know that they can’t be wrong. So, while a dexterous left hand is, technically, an oxymoron, for a minority it is a reality attracting too much attention. For centuries, prejudice and superstition have stigmatized left-handers, and several languages hint at this history by openly indicating that there is something sinister about favoring one’s left. Only recently, some left-handers may switch during their lifetime,1 and, since many were “corrected” as children, especially in former times,3,4 some may report their acquired handedness rather than the innate one. Research into handedness and mortality has been spurred, however, by the hypothesis that left-handedness is the result of an insult suffered during early development, which ultimately leads to the premature demise of left-handers.1

In this issue of Epidemiology, Ramadhani and colleagues5 contribute to this debate by presenting results from a cohort of middle-aged Dutch women followed for approximately 13 years. They report elevated colorectal cancer mortality among sinistrals compared with dextrals. Mortality for cerebrovascular disease was also increased. Breast cancer mortality, although higher, did not reach statistical significance, but incidence of premenopausal breast cancer was elevated among left-handers in this cohort.6 The analysis was careful, although the study was based on small numbers of deaths in each category.

At this point I should disclose that I am left-handed, and having successfully dodged a number of disorders, I doubt that my left hand is prematurely pulling me toward my grave. However, I am not alone in thinking that the literature on handedness suffers from a number of ills, irrespective of the putative illnesses suffered by left-handers.

Perhaps the problem from which all others follow is that it is simply too easy to include one or more questions about handedness in any questionnaire.7 That the question is easy to ask, however, does not mean that it will be answered accurately. As previously mentioned, older left-handers may classify themselves as right-handers, with the consequence that the average age of current left-handers will be younger than that of the apparent right-handers. This mechanism has led at least one deeply flawed study to conclude that left-handers’ life span was, on average, 9 years shorter than that of right-handers, thus making left-handedness as dangerous as smoking.8 Since then, several studies have addressed this topic, as reviewed by Ramadhani et al,5 with both positive and negative findings, and the question remains unresolved.

One source of confusion is that handedness is consistently determined across studies. Many assess handedness by asking hand preference in specified tasks,1,3 which translates into a degree of sidedness. Since the score usually ends up being dichotomized, this approach may be prone to posthoc manipulation of cut-offs to obtain an association.7 Others consider that the hand with which one would “naturally” write should define handedness,2 as very few are...
ambidextrous for writing, and few switch their writing hand over a lifetime unless forced to do so. Ramadhanı and colleagues\textsuperscript{2} used a longitudinal design with comparable age of left- and right-handers at the beginning of follow-up. They analyzed self-reported innate handedness, which may have been reported inaccurately but is extremely unlikely to have been differentially misclassified.

Physiologically healthy left-handers may in fact end up being less successful or less healthy than right-handers, for a variety of reasons. Perelle and Ehrman\textsuperscript{2} report a study in which teachers perceived left-handed children as having poor speech and writing, even though this was not apparent when the children were tested. Perhaps some left-handers, especially among the older generations, have been successfully persuaded that they are aberrant and slow, leading to reduced performance, lower socioeconomic status, and poorer health. It is also possible that being forced to switch to right-hand writing results in slower overall performance and, ultimately, poorer health. Social class of left- and right-handers was similar in Ramadhanı and colleagues’ study,\textsuperscript{2} although the variable used may have missed subtle differences.

One consequence of the ease with which handedness can be determined is that heterogeneity among left-handers is rarely addressed.\textsuperscript{2} The causes of handedness are unknown. Right-handedness is probably related to left-hemisphere dominance for speech. More than 90\% of right-handers have left-hemisphere language activation,\textsuperscript{3,9,10} but this is also true of approximately 75\% of left-handers, whereas 14–16\% of left-handers have symmetric activation, and the remaining right-side activation.\textsuperscript{9,11,12} Among healthy volunteers with “atypical” (nonleft) language lateralization, no significant differences were seen with respect to verbal fluency, academic achievements, and other performance indicators.\textsuperscript{13} Individuals with “weak” lateralization showed reduced disruption in language after experimentally-induced transient lesions on either side, compared with individuals with marked lateralization.\textsuperscript{14} Left-handers with aphasics may have a better prognosis than right-handers, especially if from families of left-handers,\textsuperscript{9} although the evidence is inconsistent.\textsuperscript{15}

Why do so many left-handers have left-hemisphere dominance for language and, in general, higher variability in brain lateralization? The most likely answer is that not all left-handers are born equal. Perelle and Ehrman\textsuperscript{2} suggest that there are 3 types of left-handers. First are the “pathologic” ones, ie, those who would have likely been right-handed had they not suffered some damage to their left hemisphere. These would account for the higher proportion of left-handers among those scoring highest in the verbal and mathematical portions of the Scholastic Attitude Test.\textsuperscript{2}

In general, if an insult to the left cerebral hemisphere ends up leading to left-handedness, the influx of these “forced” left-handers into the group of “normal” ones will be substantial, purely by virtue of the sheer number of those destined to be right-handed. Ramadhanı et al\textsuperscript{2} acknowledge this, although not in so many words. Too often, however the entirety of left-handers is labeled as one developmentally-disrupted entity, generally ignoring the fact that, in all the conditions putatively associated with sinistrality, dextrals retain a substantial majority.

Beyond recognizing that there are differences among left-handers, can these differences be better characterized? Hemisphere dominance cannot easily be determined in population-based studies, but attempts may be made to assess functional markers. If individuals with signs of neurologic impairment could be identified, the focus would shift from handedness to a more specific “syndrome,” which would result in the double benefit of increasing the likelihood of identifying early exposures that might be associated with a meaningful end point, while taking the pressure off left-handers in general. Another option may be to collect information about handedness in relatives and assess whether outcomes differ between sporadic versus familial left-handers. Sensitivity will be poor, however, and left-handers with no family history would still constitute a heterogeneous group.

Genetics of left-handedness is another unresolved issue. Having one left-handed parent increases one’s chance of being left-handed from 10\% to 20\%. Still, even 2 left-handed parents have about a 75\% chance of having a right-handed child. Hand discordance in monozygotic twins is about 20\%,\textsuperscript{3} although no genes for handedness have been identified. In McManus’ proposed genetic model for handedness,\textsuperscript{3} one allele, C, in double dose would give a 50\% chance of being sinistral and a 50\% chance of being dextral. Another allele, D, would always produce dextroxdextrals in double dose. The D and C alleles would act additively, and the DC genotype would thus result in 75\% dextroxdextrals and 25\% sinistrals.

Annett’s model\textsuperscript{16} proposes that inheritance of a “right-shift” gene would handicap speech-related activities in the right hemisphere and, incidentally, weaken the left hand. Those with no copy of the gene would have random and independent brain and hand asymmetry. The model implies heterozygote advantage, as homozygotes would be excessively lateralized. Annett also hypothesizes a mutation causing the gene to handicap one hemisphere at random, which could result in impairment of both hemispheres and, potentially, schizophrenia or autism.\textsuperscript{16} According to this model, reduced asymmetry will occur because neither hemisphere is impaired, or because both are.\textsuperscript{16} Over two-thirds of right-handers are “strong” dextrals, but only about one third of left-handers are “strong” sinistrals.\textsuperscript{3} This may be explained by left-handers having to live their life (however short) in a right-handers’ world. It is also compatible with the higher heterogeneity for hemisphere dominance in left-handers.
While sinistrality may share a common etiology with neurologic disorders (and, perhaps, cerebrovascular mortality), it is unclear why left-handedness would be associated with colorectal cancer mortality. This brings us to another consequence of an exposure that is too easy to assess: the frequent lack of prior hypotheses between handedness and outcome. Ramadani et al.\(^5\) do not suggest a mechanism for colorectal cancer, although they speculate that prenatal exposure to sex hormones may be at the root of the association with breast cancer, as previously suggested.\(^1^7\) That steroid hormones, particularly testosterone, may affect handedness has been hypothesized from the excess of males among left-handers. However, males are generally more vulnerable to developmental disruptions than females, and they are also more prone to some neurologic disorders, such as reading disabilities and autism-spectrum disorders.\(^1^8^,\(^1^9\)\) Boys, but not girls, whose mothers were exposed to prenatal ultrasounds were found to be more frequently left-handed,\(^2^0\) and recent findings suggest that exposure to ultrasound may interfere with neuronal migration in mouse fetuses,\(^2^1\) although these findings may not apply to humans. Coren\(^1\) reports that left-handedness is inherited more frequently from mothers than from fathers, which—if true—would fit with more males becoming left-handed as the result of nongenetic causes. While the study in this issue\(^5\) is restricted to females, the authors discuss the notion that sinistrals may have more developmental disruptions than females, and they are also more prone to some neurologic disorders, such as reading disabilities and autism-spectrum disorders.\(^1^8^,\(^1^9\)\) Boys, but not girls, whose mothers were exposed to prenatal ultrasounds were found to be more frequently left-handed,\(^2^0\) and recent findings suggest that exposure to ultrasound may interfere with neuronal migration in mouse fetuses,\(^2^1\) although these findings may not apply to humans. Coren\(^1\) reports that left-handedness is inherited more frequently from mothers than from fathers, which—if true—would fit with more males becoming left-handed as the result of nongenetic causes. While the study in this issue\(^5\) is restricted to females, the authors discuss the notion that sinistrals may have more frequently suffered developmental disruptions resulting in increased susceptibility to some diseases.\(^3\) However, the lack of a specific mechanism for colorectal cancer mortality is probably the weakest point of this report.

Ramadani and colleagues recognize that theirs might be a chance finding,\(^5\) which takes me to the final, and most problematic, item in my list of shortcomings in the literature: publication bias. The availability of handedness and the possibility of correlating it with any outcome, pose a difficult challenge to reviewers and editors. Positive studies, of good and not-so-good quality, are more likely to be published than good negative studies, which only stand a chance of being published after the positive ones have made their appearance and, sometimes, the news. This leaves the readers with the task of guessing how many unwritten or unpublished negative ones are out there for every positive finding, at least until someone produces a funnel plot that may—or may not—provide a credible estimate. Meanwhile, life goes on, even for left-handers.

**ABOUT THE AUTHOR**

Born and educated in Italy, OLGA BASSO has worked for several years at the Danish Epidemiology Science Centre in Denmark and is currently employed at the National Institute of Environmental Health Sciences in North Carolina. Her main interest is in reproductive and perinatal epidemiology, but she also has longstanding fascination with sources of bias in epidemiology.

**REFERENCES**

Statins and Cancer Risk
What Do We Know and Where Do We Go From Here?

Patricia G. Moorman* and Robert J. Hamilton†

Abstract: The relationship between statin use and cancer risk has been evaluated in numerous observational studies and as a secondary outcome in randomized controlled trials evaluating the effects of statins on cardiovascular outcomes. Although there are plausible biologic mechanisms to suggest that statins could inhibit cellular proliferation, the epidemiologic data do not show a consistent reduction in cancer risk among statin users. Despite the current lack of evidence for a chemopreventive effect, there are several methodologic considerations in the studies reported to date that prevent a definitive conclusion that statins do not reduce cancer risk. Given the widespread use of statins, continued monitoring of their risks and benefits is warranted.

(Epidemiology 2007;18: 194–196)

The lipid-lowering agents known as statins are the top therapeutic class of prescription drugs in the United States. In 2005, Lipitor® (atorvastatin; Pfizer, New York, New York) was the number one drug both in the number of prescriptions dispensed (almost 80 million) and revenue sales (more than $7 billion).¹ Data from the National Health and Nutrition Examination Surveys (NHANES) showed the prevalence of statin use tripled between surveys conducted in 1988–1994 and 1999–2002, with approximately 10% of all adults and nearly one-quarter of adults over 60 reporting statin use.² The widespread use of statins reflect their efficacy in reducing levels of total cholesterol, low-density lipoprotein cholesterol, and triglycerides and increasing high-density lipoprotein cholesterol; their increased tolerability as compared with other antihyperlipidemic agents; and the high prevalence of hyperlipidemia as a result of the epidemic of overweight and obesity in this country.

Statins have a well-established role in reducing the incidence of adverse cardiovascular outcomes, including death, myocardial infarction, atrial fibrillation, and renal dysfunction.³–6 Studies of the mechanisms of statins and observations that cardiovascular benefits are experienced even in statin users without elevated cholesterol levels led to the recognition that actions of statins extend beyond their cholesterol-lowering properties.⁷,⁸ These “pleiotropic” effects of statins are not limited to the cardiovascular system. Indeed, their effects on cell proliferation have led to the hypothesis that statins may also reduce the risk of cancer. In vitro and in vivo studies have shown statins have multiple influences on tumor growth, including inhibition of cell growth by cell cycle arrest, induction of apoptosis, suppression of angiogenesis, and other mechanisms.⁹ The idea of a widely used class of drugs such as the statins having “two-for-the-money” effects in reducing death from both cardiovascular disease and cancer certainly has great appeal. Unfortunately, the epidemiologic data accumulated to date have not yet provided convincing evidence to suggest that statins have a substantial effect on cancer incidence.

The relationship between statin use and cancer incidence has been evaluated in numerous observational studies and randomized clinical trials. Although there have been no clinical trials evaluating cancer as a primary outcome, cancer risk has been assessed in secondary analyses of many randomized controlled trials to assess the efficacy of statins for reducing cardiovascular morbidity and mortality. Several recently published meta-analyses have combined data from the nearly 3 dozen randomized controlled clinical trials of statins to evaluate cancer risk overall¹⁰–¹² or for specific cancer sites.¹³,¹⁴ Relative risks (RRs) for overall cancer incidence in individual studies ranged from approximately 0.5 to 1.8, with the most extreme values coming from trials involving fewer than 500 patients. When the combined risk for all cancers was examined, no significant association was evident in either of the 2 most recent reviews, with one reporting a RR of 0.99 (95% confidence interval = 0.94–1.04)¹¹ and the other a RR of 1.02 (0.97–1.07).¹⁰ Although randomized controlled clinical trials are generally considered the gold standard for establishing causality, the relationship between statins and cancer cannot be considered a closed issue. Despite the wealth of evidence from clinical trials, these studies had shortcomings that prevent one from making a definitive conclusion about whether statins neither increase nor decrease the risk for cancer. The randomized controlled trials were powered to detect cardiovascular outcomes, which are much more common than cancer outcomes and require shorter follow-up to accrue adequate evidence.
number of events. Among the 35 randomized controlled trials included in the meta-analysis by Bonovas et al. only 5 studies had average follow-up times of 5 years or longer, and only one had 10 years of follow-up. Not surprisingly, with the limited follow-up time, the number of cancers in any individual study tended to be fairly small. Two-thirds of the studies had fewer than 100 incident cancer cases and only one study had more than 1000. Even when combining data from multiple studies, the ability to assess the risk of individual cancer sites was limited. Another concern about making conclusions from randomized controlled trials is the possibility that the stringent inclusion criteria for these trials may limit the generalizability of the findings to a more typical, postmarketing population of statin users.

Given the limitations of the randomized controlled trials for assessing the possible role of statins in cancer etiology, data from observational studies have the potential for furthering our understanding of this relationship. Some of the largest observational studies have shown reductions in overall cancer incidence ranging from 14% to 36%, with even larger reductions in risk for specific cancer sites (eg, a 47% reduction in colorectal cancer incidence). However, other studies showed either no significant association between statin use and cancer risk or an increase in risk.

The study reported by Coogan and colleagues in this issue of Epi- demiology is an important contribution to this literature. In their hospital-based, case–control study, 4913 cases and 3900 controls recruited between 1991 and 2005 were interviewed about lifetime medication use. Major advantages of this approach include the large number of cancer cases in their study and the inclusion of individuals with long-term exposure to statins. With nearly 5000 cancer cases, the investigators were able to evaluate overall cancer risk as well as the risk for 10 individual cancer sites. The lifetime medication history allowed the investigators to consider the risk of cancer by duration of statin use. Statins were first introduced to the U.S. market in the late 1980s. Therefore, with recruitment of cases and controls through 2005, there was the potential for including individuals who had exposures for well over 10 years. As shown in the paper, the prevalence of statin use showed a steady increase in the time period during which the cases and controls were enrolled. However, even with the size of this study, the number of long-term users was fairly small, limiting the conclusions one can make about the effects of statins when used for longer durations. Given current prescribing trends and the advocacy by some experts to target ever lower cholesterol levels, it is predictable that the number of long-term users in the population will only increase.

Certainly observational studies, whether case–control studies such as the one reported by Coogan and colleagues or prospective or retrospective cohorts, have many limitations. The probability of exposure misclassification because of the reliance on self-report by participants is a very real concern. Although it is sometimes recommended that pharmacoepidemiologic studies base exposure information on medical or pharmacy records, these methods also are prone to misclassification because individuals may not take all medications that is prescribed or may obtain medications from multiple providers, making it logistically very challenging to capture all instances of medication use. Self-report of lifetime medication history, although imperfect, is a reasonable approach for drugs such as statins which are used on a chronic basis and are likely to be remembered accurately. Although the potential for errors in recall are lessened in cohort studies, a recent study by Jacobs et al pointed out that there have been important changes in statin use over time, highlighting the necessity of ascertaining exposure status at multiple points during follow-up.

The greater potential for confounding by lifestyle or medical characteristics in observational studies also has to be considered. Factors such as low physical activity, obesity, and high-fat/low-fruit-and-vegetable diets tend to cluster together, and may be related to both cancer risk and statin use. This suggests that careful assessment and analysis of confounders is critical for observational studies. Consideration of confounding by screening history is also important when assessing the association between statin use and cancer. By definition, a statin user is integrated into the medical care system. It is reasonable to assume that a patient being treated for cardiovascular disease prevention is more likely to receive preventive and screening care for other conditions. For example, a man who has been prescribed statins also may be more likely to receive prostate-specific antigen testing and be more likely to be diagnosed with prostate cancer. The failure to consider prostate cancer screening history could result in an apparent excess of prostate cancer cases among statin users and mask a protective effect of these drugs.

Given the current state of knowledge, what should be done regarding use of statins for cancer prevention? Clearly the current epidemiologic data do not warrant recommendations for statin use as a chemopreventive. Furthermore, the possibility of reducing cancer risk has to be balanced against the known adverse effects of these drugs. Although statins have a generally good safety profile, serious adverse events such as myopathies, renal dysfunction and hepatic dysfunction occur in up to 5% of patients. These adverse effects have to be given greater weight if there is any consideration of giving statins to generally healthy individuals for chemoprevention. Based on studies published to date, the risk/benefit ratio for statin use for cancer chemoprevention is not clear.

Although current data would not support recommending statin use for chemoprevention, the issue cannot be considered settled. Given the biologic plausibility of an anticancer effect of statins coupled with the limitations of published studies new studies involving adequate numbers of participants with long-term use of statins would be highly desirable. Could a randomized controlled trial be justified that is specifically designed to assess statins as chemopreventive agents, as has been suggested by some investigators? Such a randomized clinical trial would be an extremely expensive undertaking involving long-term follow-up of tens of thousands of volunteers. Based on the current conflicting evidence, it is hard to make the argument that such an undertaking is scientifically and financially justified. However, continued evaluation of the relationship between statins and cancer is critical for furthering our understanding of this relationship. Some of the studies such as the one reported by Coogan and colleagues in this issue of Epidemiology are important contributions to this literature.

© 2007 Lippincott Williams & Wilkins
cancer risk in observational studies seems prudent. In particular, further investigation is needed to determine whether effects on cancer incidence are related to dose, duration of use, extent of cholesterol lowering, or type of statin (hydrophilic versus lipophilic).

The relationship among cancer, lipid levels, and lipid-lowering drugs is another example of a situation in which the state of knowledge and the conventional wisdom have changed over time. When the first statin was introduced to the U.S. market in 1987, there were concerns that low total cholesterol was associated with increased risk for cancer.26,27 With accumulating evidence on the cellular effects of statins, the hypothesis emerged that statins could reduce cancer incidence. At present, there is no definitive answer on whether lipid-lowering drugs affect cancer risk. What is clear, however, is that any class of drugs that is being used by more than 10% of the adult population and one-quarter of the population over age 60, demands continued monitoring of its benefits and risks.

ABOUT THE AUTHORS

PATRICIA MOORMAN is an Associate Professor in the Department of Community and Family Medicine and ROBERT HAMILTON is an Associate in Research in the Department of Surgery (Urologic Surgery) at Duke University Medical Center. Dr. Moorman is an epidemiologist with longstanding interests in the etiologic relationships between medications and cancer. Dr. Hamilton is a urologist with a particular interest in the effects of statins on prostate cancer risk.

REFERENCES

22. Coogan PF, Rosenberg L, Strom BL. Statin use and the risk of ten cancers. Epidemiology. 2006;in press.
24. Kumar A, Benec C, Esserman LJ. Clinical trials are required to prove the chemopreventive worth of statins. Arch Intern Med. 2006;166:1143.
COMMENTS

Maternal Smoking and Undescended Testes: Reaching a Tipping Point

Martha Werler

Abstract: In this issue, Jensen et al (Epidemiology. 2007;18:220–225) publish new evidence linking maternal smoking with cryptorchidism. These new data make the cumulative evidence for a true association persuasive enough to be seriously considered. Possible biologic mechanisms may include endocrine disruption by cigarette smoke, although the role of endocrine disruptors in cryptorchidism more generally has been inconsistent. There are important research questions suggested by the link between mother’s smoking and cryptorchidism, including a possible role of father’s smoking, the role of genetic susceptibility, and the effects of mother’s smoking on further aspects of her son’s reproductive health.

In this issue of EPIDEMIOLOGY, Jensen and colleagues1 present fresh evidence on maternal cigarette smoking and the risk of cryptorchidism in offspring. These new data tip the balance in the direction of a true association. The authors report a positive association that joins at least 7 previous studies with similar, although less compelling, findings.2–8 Beyond adding to the tally of positive findings, the Jensen study shows trends of increasing risks with 2 measures of maternal smoking during pregnancy—increasing average number of cigarettes smoked per day and increasing average nicotine content. Further, these trends are based on prospectively collected exposure information from more than 11,000 pregnancies, with 16 to 19 years of follow-up for evidence of cryptorchidism.

How does cigarette smoking fit into a causal picture of cryptorchidism? Premature delivery and small-for-gestational age are both risk factors in that the testes normally descend relatively late in pregnancy, during the third trimester. This raises the possibility that smoking produces cryptorchidism through the developmental delay of male fetuses. However, the investigators were able to rule out this causal pathway by showing that the maternal smoking trend remained after restricting the study sample to term births with normal or heavy birth weights.

Endocrine disruption is a possible risk factor, since the testicular descent appears to be controlled by hormones. Intra-abdominal cryptorchidism, which occurs in approximately 5–10% of cases,9 represents failure of the testes to migrate to the groin in early pregnancy. Based on studies in mice, this step of testicular development is, at least in part, under control of an insulin-like factor (InsI-3), produced by Leydig cells and inhibited by estrogens. However, most undescended testes are in an inguinal position, resulting from the failure of the more complex and less understood migration into the scrotal sac. This second phase of descent is thought to be influenced by androgens.

Following the first descriptions of these 2 steps in testicular development, exogenous endocrine disruptors have been strongly suspected to be in the causal pathway of cryptorchidism.10 While there is some supporting evidence from both animals and humans, epidemiologic studies have not provided strong support.9,10 Studies of direct exposure to xenoestrogens, such as pesticides, have been hampered by the small numbers exposed. Findings on estrogen and progesterone use in pregnancy are also inconclusive, perhaps due to the timing of exposure, since most use is during early pregnancy before inguinal descent of testes occurs. Other possible markers of high estrogen levels such as nausea and vomiting in pregnancy, primiparity, and obesity, have been examined, also with mixed results.10

Does cigarette smoke fit with the endocrine disruptor theory? Environmental epidemiologists don’t typically include cigarette smoke on their lists of endocrine disruptors, although lead, cadmium, arsenic, phenols, and benzo(a)pyrene (all components of cigarette smoke) are thought to have hormonal effects. Further, altered hormone levels have been well documented in smoking women. Shorter menstrual cycle lengths and younger age at menopause are associated with cigarette smoking, raising suspicion that exposure is antiestrogenic. More specific support for endocrine disruption by cigarette smoking comes from a study by Windham and colleagues,11 who measured daily estradiol, progesterone, and follicle stimulating hormone levels across menstrual cycles and found differences between smokers and nonsmokers in all 3 hormones.

While it appears safe to say that cigarette smoking is an endocrine disruptor, we don’t know whether it has antiandrogenic effects in the fetus late in gestation, thereby inhibiting inguinal testicular descent. Certainly the toxic constituents of cigarette smoke have other adverse effects, such as oxidative stress, mutagenicity, and vasoconstriction, that might affect testicular development in the fetus.
Regardless, the findings of Jensen and colleagues should drive further research on the epidemiology and pathogenesis of cryptorchidism. Several questions are raised from their findings. First, why would the effect of father’s smoking be of similar magnitude to that of mother’s smoking? The authors raise the possibility that mothers might underreport their own smoking when their partners also smoke. This would be difficult to tease out with reported smoking levels, and should be addressed with more biologic measures of exposure.

Second, does the effect of cigarette smoking on cryptorchidism vary according to phenotype? Jensen et al found slightly lower risk estimates for boys who required surgical intervention, which is a more homogeneous subgroup than cryptorchidism as a whole (which can include boys with retractile, spontaneously descending, or absent testes or acquired cryptorchidism). One might expect the converse—a greater risk for the more narrowly defined subgroup—because misclassification is reduced. However, the effect of risk factors, including cigarette smoking, may well vary by phenotype.

Third, are hormone levels in the third trimester different according to maternal smoking status and presence of cryptorchidism at birth? Fetal (not maternal) hormone levels are most relevant, and would have to be measured during late gestation well before parturition. While the answer to this question would help elucidate the mechanism of undescended testes, and perhaps lead to identification of other risk factors, obtaining these data presents significant challenges and may be impossible.

Fourth, what genetic influences might be involved in an association between cigarette smoking and cryptorchidism? Clearly, if this association is real, it is not complete: most smoking-exposed male fetuses are born with normally descended testes. Genetically determined metabolic factors are reasonable candidates, such as polymorphisms of cytochrome P450 enzymes that are involved in both metabolizing the products of cigarette smoke and in the synthesis and breakdown of endogenous hormones.

Finally, does in utero smoking exposure play a role in the associations between cryptorchidism and increased risks of infertility and testicular cancer?12,13 If the answer is yes, the public health impact of smoking in pregnancy would be vastly extended.

The prevalence of cigarette smoking in most developed countries has decreased over the past few decades, following public health warnings. This includes decreases among women of childbearing age and pregnant women.14 Cigarette packages in the United States warn smokers of “fetal injury, premature birth and low birth weight,” and these labels, along with targeted public health efforts, have convinced approximately 40% of women who enter pregnancy as smokers to quit.14 The 10–15% of women who continue to smoke throughout pregnancy might be influenced by language about the risk of birth defects (not currently among the health warnings). Cigarette smoking already appears to increase the risk of several congenital anomalies including oral clefts and gastrochisis. If the association with cryptorchidism can be replicated in studies with biologic measures of smoking exposure, then yet another birth defect could be added to this list. Cryptorchidism is linked to decreased testicular size and sperm counts, infertility, and testicular cancer.15–17 Detailed warnings about these adverse outcomes in the unborn baby’s future might have an additional impact on both mother and father’s smoking habits during the gestation of their offspring.

ABOUT THE AUTHOR

MARTHA WERLER is Senior Epidemiologist in the Slone Epidemiology Center at Boston University and Associate Professor of Epidemiology at Boston University School of Public Health. She is a reproductive epidemiologist and has been conducting studies of maternal risk factors, such as medications, vitamins, and cigarette smoking, in relation to risks of birth defects for over 20 years.

REFERENCES

Bayesian Methods for Highly Correlated Exposure Data

Richard F. MacLehose,*† David B. Dunson,† Amy H. Herring,‡§ and Jane A. Hoppin¶

Abstract: Studies that include individuals with multiple highly correlated exposures are common in epidemiology. Because standard maximum likelihood techniques often fail to converge in such instances, hierarchical regression methods have been increasing use. Bayesian hierarchical regression places prior distributions on exposure-specific regression coefficients to stabilize estimation and incorporate prior knowledge, if available. A common parametric approach in epidemiology is to treat the prior mean and variance as fixed constants. An alternative parametric approach is to place distributions on the prior mean and variance to allow the data to help inform their values. As a more flexible semiparametric option, one can place an unknown distribution on the coefficients that simultaneously clusters exposures into groups using a Dirichlet process prior. We also present a semiparametric model with a variable-selection prior to allow clustering of coefficients at 0. We compare these 4 hierarchical regression methods and demonstrate their application in an example estimating the association of herbicides with retinal degeneration among wives of pesticide applicators.

Epidemiology 2007;18: 199 –207)

Highly correlated exposures are ubiquitous in epidemiologic research and may arise as the result of an association between the measured exposures and one or more latent factors. For example, pesticide exposures for farm workers may be highly correlated because individuals apply multiple pesticides in a year, with choice of pesticide influenced by type of crop and pest.1,2 To depict this correlated exposure problem, let \( x_1, \ldots, x_k \) denote the levels of \( k \) different exposure variables that are highly correlated due to an unmeasured variable or variables, and let \( y \) denote the outcome. Researchers will generally be interested in estimating effect measures \( \beta_1, \ldots, \beta_k \) for exposures \( x_1, \ldots, x_k \). Hence, a common strategy is to fit the logistic regression model:

\[
\logit \{ \Pr(y_i = 1|x_{i1}, \ldots, x_{ik}) \} = \alpha_0 + \beta_{1x_{i1}} + \ldots + \beta_{kx_{ik}}.
\]

Unfortunately, maximum likelihood estimation of the model in expression (1) can fail to converge when predictors are highly correlated, and estimated coefficients may be unstable even when convergence is achieved.3

This problem has led many epidemiologists to fit logistic regression models incorporating one exposure variable at a time. However, the other exposure variables may be confounders and, if so, must be included to assess the causal effect of any specific exposure.4 Another commonly used strategy is to collapse the specific exposure information into summaries, such as a sum across chemicals in a class or an ever/never indicator. Unfortunately, this strategy results in a loss of information, does not allow inferences on effects of specific exposures, and can be sensitive to the chosen summary measure.

The problems associated with performing maximum likelihood estimation on correlated data have resulted in an increased use of hierarchical models.5 Ordinary regression models treat the outcome as a random variable, dependent on parameters. For example, in expression (1), \( y_i \) is a random variable that depends on the parameters \( \alpha_0 \) and \( \beta_1 \ldots \beta_k \). Hierarchical regression extends ordinary regression by also treating parameters as random variables depending on further coefficients through a prior distribution. Estimates obtained through hierarchical regression are shrinkage estimates in the sense that they are moved away from the asymptotically unbiased maximum likelihood estimate (MLE) and toward the center of the prior distribution. Shrinkage estimators are advantageous in 2 ways: they often have smaller frequentist mean squared error (MSE) and they represent incorporation of prior knowledge in the Bayesian sense.6 Such hierarchical models help circumvent problems associated with MLE. Namely, hierarchical models can estimate effects with lower MSE, even in the presence of high correlation.3,7

We discuss 4 Bayesian hierarchical models: 2 parametric models (P1 and P2) and 2 semiparametric models (SP1 and SP2). These 4 models differ in how their prior distribution is specified. The most common Bayesian hierarchical model found in epidemiologic research is the semi-Bayes model,5,8–13 which we refer to as model P1 (ie, the first parametric model). A typical prior distribution for \( \beta_j \) (where \( j \) indexes the \( k \) coefficients in expression 1) is \( N(\mu, \phi_j) \), where \( \mu \) characterizes the investigator’s prior knowledge about the
true value of β_j and φ^2 is the uncertainty regarding that value. Values for μ and φ^2 are chosen based on substantive knowledge. The amount that the estimated effects are shrunk away from the MLEs and toward the prior mean is determined by the prior variance, φ^2. A large prior variance indicates greater uncertainty about the effect size and causes less shrinkage.

Consider a model such as expression (1) in which 20 coefficients are estimated and each has a N(0, φ^2) prior. Prior knowledge may exist about the variability of the estimates, but the data also contain information about that variability, with a simplistic estimate being the variance of the 20 MLEs about the prior mean. Model P1 incorporates prior knowledge by treating φ^2 as known, but it ignores information regarding the variability of the coefficients that is contained in the observed data. Thus, model P1 has fixed shrinkage regardless of the support for the prior distribution provided by the data.

Consider, instead, a model that treats these prior parameters as random variables in turn having their own prior distributions (model P2). Unlike model P1, which has fixed shrinkage (because φ^2 is constant), model P2 estimates φ^2 by combining the observed data with prior knowledge about φ^2. This allows the amount of shrinkage to vary depending on how well the data support the prior distribution. If the data lend some support to the prior distribution, model P2 can provide greater shrinkage than model P1. If the data lend little support to the prior distribution, model P2 will result in less shrinkage. In the discussion so far, all coefficients have been shrunk toward a common mean; however, it is straightforward to allow coefficients to be grouped into classes with each set of coefficients shrunk toward separate class-specific means.

Models P1 and P2 have potential disadvantages. A normally distributed prior is commonly assumed for historical reasons and computational convenience; however, results may be sensitive to this assumption. Second, for these methods to shrink estimates towards multiple prior means, the coefficients must be specified into classes (eg, if the coefficients are the effects of different pesticides, they could be classified as fungicides or fumigants to allow coefficients in those classes to be shrunk toward different means). However, it may be impossible to specify which effects should be grouped into which classes, or even how many classes there should be. A method that allows the data to guide the clustering of coefficients into classes would be preferable. To accomplish this, we place a Dirichlet process prior (DPP) on the distribution of the coefficients.14–16 A DPP allows researchers to specify their prior knowledge as being “similar” to a known parametric distribution (such as the normal) while remaining flexible enough to allow for substantial deviations from that distribution. Additionally, the DPP attempts to cluster coefficients into groups based on effect size. Coefficients are clustered together probabilistically (soft clustering) rather than with certainty (hard clustering) and this feature of DPPs can offer dramatic improvements in effect estimation. We will refer to this semiparametric model with DPP priors as model SP1.

In epidemiologic studies, some exposures will typically have virtually no effect, in which case they cannot confound the effect of any other exposure and we might prefer to exclude them from the model. Variable-selection techniques in the epidemiologic literature are limited, generally relying on backward or forward selection strategies that increase the type I error rate.17–19 However, there has been an increasing focus on variable-selection methods (implemented through variable-selection priors) in the statistics literature based on the advent of microarray technology.20,21

To account for the opportunity that some β_j = 0, we propose a mixture prior that allows an unknown subset of the predictors to have zero coefficients.22,23 A coefficient is implicitly removed from the model when β_j = 0, a probability we estimate by combining our prior knowledge of a null effect with the observed data. When using a DPP for the coefficients, the exposures are clustered into groups. By using this mixture prior, we also allow a cluster of exposures that has coefficients equal to zero. Adopting this prior distribution in the DPP to perform simultaneous variable selection and clustering has been shown to have excellent properties.24 We refer to the semi-parametric model with clustering of coefficients at zero as model SP2.

**Parametric Models**

Both parametric models (P1 and P2) have been discussed in much greater detail elsewhere.5,6,11,12,25,26 Here, we illustrate some of their properties in the simple setting of an ordinary linear regression model in which centered covariates x_1, . . . , x_k are regressed on an outcome y. For ease of presentation, we assume the linear model has a known error term, σ^2, no intercept, and that the covariates are orthogonal (ie, they are not correlated); however, the results are generalizable to nonorthogonal situations.

As mentioned previously, model P1 incorporates information on β_j through a prior distribution. A specific formulation for this model is:

$$y_j | β_j \sim N \left( \sum_{j=1}^{k} β_j x_{j}, \sigma^2 \right)$$

$$[β_j] \sim N \left( η_j, φ^2 \right)$$  \hspace{1cm} (2)

where the prior mean, η_j, incorporates prior evidence regarding the size of the effect for the jth coefficient and the x_j may be standardized so they are all on the same scale. Prior scientific knowledge may indicate that all coefficients have the same prior distribution, that some coefficients have one prior distribution while others have a different prior distribution, or that each coefficient has its own prior distribution. For example, if β_1, . . . , β_k are the effects of pesticides on retinal degeneration, one could assume that the effects of all pesticides are the same (eg, they all belong to the same class and have a common prior distribution), that the effect varies over different functional groups of pesticides (eg, they could be grouped into classes such as fungicide or fumigant, with each class having a different prior distribution), or that each
pesticide has a different prior distribution. Indicator variables, \(z_{jk}\) denoting a pesticide class can be introduced into the prior distribution by allowing \(\eta_j = \sum_{k=1}^{p} \theta_j z_{jk}\). However, these classes need not be mutually exclusive and more complicated prior specifications can be included where biologically relevant. The prior variance \(\phi_j^2\) represents the uncertainty that \(\beta_j = \eta_j\). The prior variance could be specified from a meta-analysis or could be calculated by choosing a range within which the researcher believes 95% of effect estimates on this topic would lie. Solving the standard confidence interval formula for the variance term allows the researcher to specify the prior variance. The lack of a prior distribution on \(\eta_j\) or \(\phi_j^2\) is the distinguishing feature of model P1.

The posterior distribution (ie, the distribution that results when the prior distribution for \(\beta_j\) is updated with the observed data) for \(\beta_j\) is given by:

\[
[\beta_j|Data] \sim N \left( \frac{\eta_j/\phi_j^2 + \sum x_{ij}y_j/\sigma_j^2}{1/\phi_j^2 + \sum x_{ij}^2/\sigma_j^2} \right)
\]

The posterior mean is an average of the prior mean (\(\eta_j\)) and the maximum likelihood estimate (\(\sum x_{ij}y_j/\sum x_{ij}^2\)), inversely weighted by their respective variances, \(\phi_j^2\) and \(\sigma_j^2/\sum x_{ij}^2\). This is the essence of a shrinkage estimator: the posterior distribution of \(\beta_j\) is shrunk towards its prior distribution. As the number of observations increase, the posterior distribution is weighted more heavily toward the observed data. With orthogonal data of moderate size, the observed data will quickly overwhelm anything but the strongest priors (ie, those with very small \(\phi_j^2\)), and estimates obtained from these parametric Bayesian models will be similar to the MLE. For concreteness, we generated a small (n = 50) dataset with 5 orthogonal covariates, none of which have an effect. We estimate \(\beta_1, \ldots, \beta_5\) using MLE and using model P1 in expression 2 with \(k = 5\) and \(\eta_j = 0\) for all \(j\). Figure 1 shows the resulting distribution of the MLE of \(\beta_j\), as well as the Bayesian estimate with \(\phi_j^2 = 0.5, 1.0,\) and 2.0. Note that, on average, the MLE will be unbiased but in this single sample the results are far from the truth. The amount of shrinkage in the hierarchical models is a function of the prior variance: as the prior variance decreases (representing increasing certainty about the effect of \(\beta_j\)), the posterior distribution shrinks toward the prior mean.

Because \(\phi_j^2\) is so vital to model P1, it is wise to vary it in sensitivity analyses and see how estimates change with different plausible values. In Figure 1, for example, \(\phi_j^2 = 1.0\) may have been the best guess of the variance of \(\beta_j\) with sensitivity analyses conducted for \(\phi_j^2 = 2.0\) and \(\phi_j^2 = 0.5\). However, although there may be uncertainty regarding the prior variance (leading to sensitivity analyses), estimates from model P1 cannot account for this uncertainty.

Model P2 explicitly accounts for uncertainty in the prior variance by placing a prior distribution on \(\phi_j^2\), resulting in estimates that are averaged over plausible values of \(\phi_j^2\).

Unlike the fixed shrinkage of model P1, model P2 adapts the shrinkage of \(\beta_j\) based on the observed variability of \(\beta_{j1}, \ldots, \beta_{jk}\) from their prior mean. Additionally, when the prior mean is a function of covariates (eg, \(\eta_j = \sum \theta_j z_{jk}\)), substantive information may exist for the effect of those variables and a prior distribution can be placed on those parameters. A typical specification for model P2 is:

\[
[y_j|\beta_j] \sim N \left( \sum_{j=1}^{k} \beta_j x_{ij}, \sigma_j^2 \right)
\]

\[
[\beta_j|\theta, \phi_j^2] \sim N \left( \sum_{j=1}^{p} \theta_j z_{ij}, \phi_j^2 \right)
\]

\[
[\theta_j] \sim N (\mu_j, \omega_j^2)
\]

\[
[\phi_j^2] \sim IG (\alpha_{ij}, \alpha_{ij})
\]

Here, \(\theta_j\) is the effect of a \(z_{ij}\) covariate and its prior mean, \(\mu_j\) is the prior knowledge regarding the size of \(\theta_j\)‘s effect; the prior variance \(\omega_j^2\) represents uncertainty in that effect. The prior distribution for the \(\phi_j^2\) is chosen as an inverse gamma (IG) distribution with parameters \(\alpha_{ij}^1\) and \(\alpha_{ij}^2\). The inverse gamma distribution is a common choice for the prior distribution of a variance term because of its flexibility and computational convenience. The prior mean of \(\phi_j^2\) is \(\alpha_{ij}^1/(\alpha_{ij}^1 - 1)\) and its variance is \(\alpha_{ij}^2/(\alpha_{ij}^1 - 1)^2(\alpha_{ij}^2 - 2)\). Model P1 is a special case of model P2 in which the variance of \(\theta_j\) and \(\phi_j^2\) goes to zero. In choosing values of \(\alpha_1\) and \(\alpha_2\) for an analysis, we suggest specifying a most likely value of \(\phi_j^2\) (call this \(E_{\phi_j^2}\)) and a value for the variance of \(\phi\) (call this \(V_{\phi_j^2}\)) such that 95% of the reasonable \(\phi_j^2\) values would fall within the 95% confidence interval (CI) for \(IG(E_{\phi_j^2}, V_{\phi_j^2})\). Solving the mean and variance equations for \(\alpha_1\) and \(\alpha_2\) gives: \(\alpha_1 = E_{\phi_j^2}/V_{\phi_j^2} + 2\)
and $\alpha_2 = E[p_0 \phi_0] + E[\phi_0]$. It is useful to plot a large number ($n = 10,000$) of samples from the prior to ensure that the shape of distribution conforms to prior knowledge.

The full conditional posterior distributions for the parameters in model P2, assuming $\phi^2$ is the same for all $\beta_j$ (for simplicity) are:

$$[\beta_j|Data, \sigma^2_j, \theta_j, \phi^2] \sim N \left( \frac{\sum_i x_{ij} \phi^2 + \sum x_{ij} \theta_j}{1/\phi^2 + \sum x^2_{ij}/\sigma^2}, \frac{1}{1/\phi^2 + \sum x^2_{ij}/\sigma^2} \right)$$  (5)

$$[\theta_j|Data, \beta_j, \phi^2] \sim N \left( \frac{\mu_j/\omega_j^2 + \sum z_{ij} \beta_j / \phi^2}{1/\omega_j^2 + \sum z^2_{ij} / \phi^2}, \frac{1}{1/\omega_j^2 + \sum z^2_{ij} / \phi^2} \right)$$  (6)

$$[\phi^2|Data, \beta_j, \theta_j] \sim IG \left( \alpha_1 + p/2, \alpha_2 + \frac{\sum (\beta_j - \sum z_{ij} \theta_j)^2}{2} \right)$$  (7)

The conditional distribution of $\phi^2$ in expression (7) is of particular interest. The adaptive shrinkage properties of model P2 are apparent from the $\sum (\beta_j - \sum z_{ij} \theta_j)^2$ term, that represents the variation of the $\beta_j$ from their prior mean. As the variance of the parameters increases, the mean of $\phi^2$ also increases and when the variance decreases, the mean of $\phi^2$ decreases. Thus, if the data indicate that our prior specification of $\phi^2$ is too small, the posterior mean of $\phi^2$ is increased to reflect this. Because $\phi^2$ determines the amount of shrinkage, $\beta_j$ will be shrunk to a lesser extent. The converse is also true; when the data show little variability of the estimates from the prior mean, the posterior estimate of $\phi^2$ will decrease and cause greater shrinkage of $\beta_j$ to their prior distribution. This adaptive shrinkage is a potential improvement over model P1 that has a constant amount of shrinkage, regardless of the variability of the $\beta_j$ from the prior mean that is observed in the data. Model P2 also allows inferences to be more data-driven and less sensitive to the prior specification of $\mu$ and $\phi^2$.

The distribution of $\beta_j$ in expression (5) is similar to the distribution in expression (3). However, the distribution from model P1 is conditional on known values, while the distribution from model P2 is conditional on random variables ($\phi^2$ and $\theta_j$). To average the distribution of $\beta_j$ over these random variables, we use Gibbs sampling (a type of Markov Chain Monte Carlo) that proceeds by iteratively drawing parameter values from the full conditional distributions in expressions (5, 6) and (7), given the value of the other random variables from the previous steps of the Gibbs sampler. After running the Gibbs sampler for a large number of iterations and discarding some initial number of iterations to allow for a burn-in period, the mean and variance of $\beta_j$ in the remaining samples are the mean and variance of the marginal posterior distribution of interest. For more information regarding burn-in period and convergence, consult Gelman et al. We also note that these algorithms (which can be implemented in programs such as WinBUGS [MRC Biostatistics Unit, Cambridge, UK]) generate the exact posterior distribution of the coefficients that is useful in small datasets. This result is an improvement over previous methods proposed for fitting model P1 that rely on asymptotic approximations.

We analyze, under model P2, the dataset we previously examined for the model P1. The prior mean for all $\beta_j$ is zero and the parameters for the prior variance, $\phi^2$, are $\alpha_1 = 1$ and $\alpha_2 = 1$. We ran a Gibbs sampling algorithm for 50000 iterations and excluded the first 5000 iterations as a burn-in period. The marginal posterior distributions of $\beta_1$ and $\phi^2$ are presented in Figure 2. The mean of $\beta_1 = -0.51$, which is between the mean of the estimates from model P1 under the assumption of a fixed $\phi^2 = 1$ ($\beta_1 = -0.56$) and $\phi^2 = 0.5$ ($\beta_1 = -0.43$). Although the mean of the prior variance was 1 in the model P2, $\beta_1 \ldots \beta_5$ exhibited less variability than the prior indicated, and the posterior mean of $\phi^2 (0.87)$ decreased to reflect this additional information. Thus, by incorporating information on $\phi^2$ that is contained in the data, we adaptively allow greater shrinkage of $\beta_1$ towards its prior mean.

Although we have focused on linear regression with orthogonal data, the results can be generalized to correlated predictors and logistic regression. It is only for computational convenience that we have focused on linear models here. We implement logistic hierarchical models in simulations and the applied example presented later in this paper.

**Semiparametric Models**

As we demonstrate (Appendix I, available with the online version of this article), models P1 and P2 can offer a distinct improvement over MLE. However, results of these
models may be sensitive to the assumed prior distribution of $\beta$, and a nonparametric prior may be preferable. Further, when sufficient prior information exists, coefficients may be grouped into classes by incorporating second level coefficients; however, in many epidemiologic applications such prior knowledge may not exist. Instead, we explore a procedure that allows coefficients to be grouped into clusters based on similarity of effect sizes before shrinking them toward a prior distribution.

In Bayesian nonparametric inference, a common method to limit the dependence of a parameter on a prior distribution is to let the prior distribution be random. In the previous section we assumed $\beta_j \sim N(\mu, \sigma^2)$. Instead, we could specify $\beta_j \sim D$, where $D$ is a random distribution. Because $D$ is random, we place a prior distribution on it; in this case we choose a Dirichlet process prior, $D \sim DPP(\lambda D_0)$, where $D_0$ is the base distribution (such as a normal distribution) and $\lambda$ is a precision parameter determining how closely $D$ follows $D_0$. As $\lambda$ increases, $D$ converges to $D_0$, and the nonparametric approach reduces to the parametric models of the previous section. Smaller values of $\lambda$ indicate less certainty that $\beta_j \sim D_0$. Figure 3 presents 2 realizations of $DPP(\lambda D_0)$ with $D_0 = N(0, 1)$ and $\lambda$ equal to either 1 or 100. The larger value of $\lambda$ yields a distribution that resembles the base distribution, while the sample with $\lambda = 1$ shows no similarity to the $D_0$.

A feature of the 2 distributions shown in Figure 3 is their discrete nature. Rather than being continuous, like the base distribution, a draw from a DPP is discrete, implying that any 2 (or more) coefficients have a nonzero probability of being clustered together and having the same effect size. The scale of the predictor may be important, and the predictors could potentially be rescaled to allow greater similarity among coefficients. The clustering feature can be seen more clearly through the (identical) representation of the DPP as a mixture distribution: $\beta_j \sim \omega_0 D_0 + \omega_1 \sum_i \delta_{\beta_i}$, where $\omega_0$ and $\omega_1$ are weights determined by $\lambda$. The term $\delta_{\beta_i}$ indicates that, with probability $\omega_1$, $\beta_j$ is clustered with coefficient $\beta_i$. The posterior probability of clustering coefficients depends on $\lambda$ (smaller values of $\lambda$ favor clustering) and the similarity of the magnitude of those coefficients (increased similarity favors clustering).

Consider 2 predictors, $x_{im}$ and $x_{in}$, with effects $\beta_m$ and $\beta_n$ which follow some unknown distribution $D$ that is assigned a DPP. This model estimates, based on prior knowledge and information in the data, a probability $p_{mn}$ that $\beta_m = \beta_n$. In the extreme (and unlikely) case where $p_{mn} = 1$, coefficients $x_{im}$ and $x_{in}$ are estimating parameters with the same value ($\beta_m = \beta_n$). That is, the data contain twice as much information regarding the common effect, resulting in more precise effect estimates as well as less shrinkage toward the prior distribution. At the other extreme, if $p_{mn} = 0$, the 2 coefficients do not aid in each other’s estimation. More commonly, $p_{mn}$ will be between 0 and 1, allowing $\beta_m$ and $\beta_n$ to add some information to one another’s estimation. This will result in different posterior distributions for the 2 coefficients that can have lower MSE than model P1 or P2 (see Appendix I). In the sense that model P1 allowed for constant shrinkage of all coefficients toward the prior mean, and model P2 allowed for adaptive shrinkage of all coefficients toward the prior mean, the semiparametric models (SP1 and SP2) allow individual coefficients to be adaptively shrunk toward the prior mean to different extents. The more likely coefficients are to be clustered together, the more information there is in the data regarding their common effect, and the less impact the prior specification will have.

This model is semiparametric because the distribution of the outcome, $y_i$, is parametric, whereas the distribution of $\beta_j$ is nonparametric. The first semi-parametric

---

**FIGURE 3.** Two simulations from $DPP(\lambda D_0)$ with $\lambda = 1$ (left) and $\lambda = 100$ (right). $D_0 \equiv N(0, 1)$. 

© 2007 Lippincott Williams & Wilkins
Methods are available to estimate this parameter as well. As serves the same function as placing a parameter on noninformative values are chosen for than relying solely on prior knowledge. Generally, relatively model P1; it allows the data to help guide inference rather than empirical Bayes approaches favoring the same data used by Kirrane et al in their analysis of those for whom application to study of pesticides and retinal degeneration among the farmers wives. Wives filled out a questionnaire with between pesticide exposure and retinal degeneration among elsewhere.1 Kirrane et al2 recently examined the association between 1993 and 1997, has been described in more detail elsewhere. In many instances it may be useful to exclude variables that have no effect on the outcome or there may be prior substantive knowledge that the exposure has no effect. In either case, modification of the base distribution in expression 8 allows a variable-selection prior distribution to be incorporated in a DPP model. Following the approach of Dunson et al,34 we specify a second semi-parametric model (SP2):

\[
y_i \sim N \left( \sum_{j=1}^{k} \beta_j x_{ij}, \sigma^2 \right)
\]

\[
\beta_j \sim D
\]

\[
D \sim DP(\lambda D_0)
\]

\[
D_0 = N(\mu, \phi^2)
\]

\[
\lambda \sim G(a, b)
\]

\[
\pi \sim \text{beta}(c, d)
\]

\[
\phi^2 \sim IG(\alpha_1, \alpha_2)
\]

where \(G\) is a gamma distribution with mean \(ab\) and variance \(ab^2\). Placing a prior distribution on the precision parameter, \(\lambda\), serves the same function as placing a parameter on \(\phi^2\) in model P1; it allows the data to help guide inference rather than relying solely on prior knowledge. Generally, relatively noninformative values are chosen for \(a\) and \(b\), such as \(a = 1\), \(b = 1\) or \(a = 0.1\), \(b = 0.1\). However, empirical Bayes methods are available to estimate this parameter as well.31 As with the model P2, estimating these parameters requires a Gibbs sampling algorithm.

In many instances it may be useful to exclude variables that have no effect on the outcome or there may be prior substantive knowledge that the exposure has no effect. In either case, modification of the base distribution \(D_0\) in expression 8 allows a variable-selection prior distribution to be incorporated in a DPP model. Following the approach of Dunson et al,34 we specify a second semi-parametric model (SP2):

\[
y_i \sim N \left( \sum_{j=1}^{k} \beta_j x_{ij}, \sigma^2 \right)
\]

\[
\beta_j \sim D
\]

\[
D \sim DP(\lambda D_0)
\]

\[
D_0 = \pi \delta_0 + (1 - \pi) N(\mu, \phi^2)
\]

\[
\lambda \sim G(a, b)
\]

\[
\pi \sim \text{beta}(c, d)
\]

\[
\phi^2 \sim IG(\alpha_1, \alpha_2)
\]

where \(\delta_0\) indicates a point mass at the value 0. The base distribution has a value of 0 with probability \(\pi\), and distribution \(N(\mu, \phi^2)\) with probability \(1 - \pi\). This simple modification to the base distribution allows \(\beta_j\) to be equal to 0, in which case it is effectively removed from the model. This exclusion can help increase the precision of estimates, particularly in the presence of highly correlated variables or in small datasets. An important feature of allowing coefficients to be equal to 0 is that it allows for easy testing of a point hypothesis, such as \(\beta_j = 0\). For instance, if a Gibbs sampling algorithm is run for \(R\) iterations and \(\beta_j = 0\) for \(r\) of those iterations, the posterior probability that \(\beta_j = 0\) is \(r/R\).

When \(\pi = 0\), model SP2 reduces to the model SP1. The coefficient \(\pi\) is given a \(\text{beta}(c, d)\) distribution to allow the data to inform the probability that a coefficient is zero. Elicitation of \(c\) and \(d\) can proceed by specifying the expected probability, \(E_{\pi}\), that a randomly selected coefficient is 0 and the variance surrounding that estimate, \(V_{\pi}\). Solving the equations for the mean and variance of the beta distribution:

\[
c = \frac{E_{\pi} - E_{\pi}^3}{V_{\pi}} - E_{\pi}
\]

\[
d = \frac{E_{\pi}(E_{\pi} - 1)^2}{V_{\pi}} + E_{\pi} - 1.
\]

Example: Application to Study of Pesticides and Retinal Degeneration

The Agricultural Health Study, which enrolled farmers who applied for pesticide licenses in Iowa or North Carolina between 1993 and 1997, has been described in more detail elsewhere.1 Kirrane et al2 recently examined the association between pesticide exposure and retinal degeneration among the farmers wives. Wives filled out a questionnaire with information on their medical and pesticide history. We analyzed the same data used by Kirrane et al in their analysis (31,173 women, 281 of whom experienced retinal degenera-

### Table 1. Hierarchical Models Used to Analyze Agricultural Health Study Data on Herbicides and Retinal Degeneration in the Wives of Pesticide Applicators, North Carolina and Iowa, 1993–1997

<table>
<thead>
<tr>
<th>P1</th>
<th>P2</th>
<th>SP1</th>
<th>SP2</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta_j \sim N(0, 0.35))</td>
<td>(\beta_j \sim N(0, \phi^2))</td>
<td>(\beta_j \sim D)</td>
<td>(\beta_j \sim D)</td>
</tr>
<tr>
<td>(\phi^2 \sim IG(2.13, 0.40))</td>
<td></td>
<td>(D \sim DP(\lambda D_0))</td>
<td>(D \sim DP(\lambda D_0))</td>
</tr>
<tr>
<td>(D_0 = N(0, \phi^2))</td>
<td></td>
<td>(D_0 = \pi \delta_0 + (1 - \pi) N(0, \phi^2))</td>
<td>(D_0 = \pi \delta_0 + (1 - \pi) N(0, \phi^2))</td>
</tr>
<tr>
<td>(\lambda \sim G(1, 1))</td>
<td></td>
<td>(\lambda \sim G(1, 1))</td>
<td>(\lambda \sim G(1, 1))</td>
</tr>
<tr>
<td>(\phi^2 \sim IG(2.13, 0.40))</td>
<td></td>
<td>(\phi^2 \sim IG(2.13, 0.40))</td>
<td>(\pi \sim \text{beta}(1.5, 1.5))</td>
</tr>
</tbody>
</table>
tion), but we limit our analysis to herbicides, of which there are 18 unique chemicals. These 18 chemicals exhibited a wide range of correlation, from 0.06 to 0.58. Table 1 shows the 4 hierarchical models used to analyze the data. Prior parameter values are based on prior knowledge and are similar to those used in Kirrane et al. There is little evidence of an effect of herbicides on retinal degeneration, so we center our prior distributions at OR = 1.0. Gibbs sampling algorithms were programmed in Matlab (The Mathworks, Natick, MA) and run for 60,000 iterations, with the initial 5,000 excluded as a burn-in period.

To help illustrate the 4 hierarchical models, we present representations of the prior distributions for the effect of the herbicide imazethapyr ($\beta_1$) in Figure 4. Since the prior distributions for models P2, SP1 and SP2 depend on random variables, we evaluate their prior distributions at the posterior mean of all other random variables. The prior distribution for model P2 is of particular interest. The large spike observed at 0 is the posterior probability that $\beta_1 = 0$ with probability 0.68 and $\lambda = 1.5$ imply that $\beta_1$ is distributed according to $N(0, 0.18)$ with probability 0.03 or set equal to one of $\beta_2 \ldots \beta_{18}$ with probability 0.29 or set equal to 0 with probability 0.68.

The results of the models are presented in Table 2. Figure 5 shows the posterior distribution of the effect of imazethapyr from the 4 hierarchical models. Model P1 estimated an effect of imazethapyr that was no longer statistically significant but still markedly elevated (OR = 1.7; 95% CI = 0.8 – 3.6). Models P2, SP1, and SP2 were all largely in agreement, indicating little evidence of effect of imazethapyr on retinal degeneration. The distribution of $\beta_1$ estimated through model SP2 is of particular interest. The large spike observed at 0 is the posterior probability that $\beta_1 = 0$ ($P = 0.63$). Also of interest in the posterior distribution from model SP2 is the fact that the most likely non-null effect is virtually than indicated by the prior; this value implies that with probability 0.2, $\beta_1$ is distributed according to $N(0, 0.17)$, and with probability 0.8, $\beta_1$ is assigned the value of one of the other coefficients, $\beta_2 \ldots \beta_{18}$. The prior distribution for model SP2 is similar to model SP1, except for a large point mass at 0. The posterior mean of $\pi = 0.68$ and $\lambda = 1.5$ imply that $\beta_1$ is distributed according to $N(0, 0.18)$ with probability 0.03 or set equal to one of $\beta_2 \ldots \beta_{18}$ with probability 0.29 or set equal to 0 with probability 0.68.

DISCUSSION

Although highly correlated data are common in epidemiology, standard analyses can produce extremely imprecise confidence intervals or fail to converge altogether. We examined 4 Bayesian hierarchical models that perform well in this context. These models may have broad use in other areas, as well, such as in estimating models with a large number of predictors.

When deciding which of the 4 models to use in an analysis, consideration should be given to the properties of...
each model as well as the computational skill required to implement them. The 2 parametric models (P1 and P2) are the easiest computationally. Either model can be implemented in WinBUGS, using the code we provide in Appendix II. The advantages of model P2 may justify its use in preference to model P1. Model P1 assumes a fixed prior variance, whereas model P2 updates the prior variance based on the observed data. This “Bayesian learning” allows for adaptive shrinkage and makes estimates more data-driven and less sensitive to prior specification. However, as the sample size increases, the difference between model P1 and P2 (and ML) tends to decrease.

Although more computationally intensive than the parametric models, the 2 semiparametric models presented here have very desirable properties in many situations. These models may be particularly useful when the researcher is unaware how to group coefficients. When some coefficients have similar true values, the semiparametric models can decrease MSE by aggregating data within clusters. Indeed, even if the true values of the coefficients are not exactly identical, “soft clustering” can still reduce MSE (see Appendix I). However, as the probability of clustering coefficients increases, models SP1 and SP2 can perform remarkably well. The decision whether to implement model SP1 or SP2 should be made on substantive grounds. When researchers have a high prior probability that many of the effects in question may be zero, the selection prior in model SP2 can help estimation. Model SP2 may be particularly useful in situations where hypothesis testing is required. However, when the true value of most coefficients is zero and only a few coefficients are nonzero (but still close to zero), model SP2 performs slightly worse than model SP1.

In summary, the challenges of analyzing highly correlated data can be greatly diminished using the Bayesian framework. The 2 parametric and 2 semiparametric models we examine in this paper provide useful alternatives to

<table>
<thead>
<tr>
<th>Herbicide</th>
<th>MLE OR (95% CI)</th>
<th>P1 OR (95% CI)</th>
<th>P2 OR (95% CI)</th>
<th>SP1 OR (95% CI)</th>
<th>SP2 OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imazethapyr</td>
<td>2.6 (1.0–6.3)</td>
<td>1.7 (0.8–3.6)</td>
<td>1.3 (0.8–2.5)</td>
<td>1.1 (0.5–3.8)</td>
<td>1.1 (0.8–2.3)</td>
</tr>
<tr>
<td>Chlorimuronethyl</td>
<td>1.9 (0.7–5.0)</td>
<td>1.4 (0.6–3.0)</td>
<td>1.2 (0.7–2.1)</td>
<td>1.0 (0.4–2.5)</td>
<td>1.0 (0.7–1.6)</td>
</tr>
<tr>
<td>Alachlor</td>
<td>1.4 (0.6–3.1)</td>
<td>1.2 (0.6–2.2)</td>
<td>1.1 (0.6–1.8)</td>
<td>0.9 (0.3–1.7)</td>
<td>1.0 (0.6–1.3)</td>
</tr>
<tr>
<td>Petroleum oil</td>
<td>1.4 (0.7–2.9)</td>
<td>1.3 (0.7–2.4)</td>
<td>1.2 (0.7–1.9)</td>
<td>1.0 (0.4–1.9)</td>
<td>1.0 (0.7–1.4)</td>
</tr>
<tr>
<td>2,4,5-TP</td>
<td>1.3 (0.1–11.2)</td>
<td>1.0 (0.3–2.6)</td>
<td>1.0 (0.5–1.9)</td>
<td>1.0 (0.4–2.3)</td>
<td>1.0 (0.6–1.6)</td>
</tr>
<tr>
<td>2,4-D</td>
<td>1.3 (0.8–1.9)</td>
<td>1.2 (0.8–1.8)</td>
<td>1.2 (0.8–1.7)</td>
<td>1.0 (0.5–1.7)</td>
<td>1.0 (0.6–1.4)</td>
</tr>
<tr>
<td>Butylate</td>
<td>1.1 (0.3–3.9)</td>
<td>1.0 (0.4–2.3)</td>
<td>1.0 (0.5–1.8)</td>
<td>1.0 (0.4–2.0)</td>
<td>1.0 (0.6–1.5)</td>
</tr>
<tr>
<td>Glyphosate</td>
<td>1.1 (0.8–1.5)</td>
<td>1.0 (0.8–1.4)</td>
<td>1.1 (0.8–1.4)</td>
<td>0.9 (0.4–1.4)</td>
<td>1.0 (0.6–1.3)</td>
</tr>
<tr>
<td>Dicamba</td>
<td>1.0 (0.4–2.2)</td>
<td>1.0 (0.5–1.9)</td>
<td>1.0 (0.6–1.7)</td>
<td>0.9 (0.4–1.6)</td>
<td>1.0 (0.6–1.4)</td>
</tr>
<tr>
<td>Trifluralin</td>
<td>1.0 (0.5–2.1)</td>
<td>1.0 (0.5–1.8)</td>
<td>1.0 (0.6–1.7)</td>
<td>0.9 (0.4–1.9)</td>
<td>1.0 (0.6–1.5)</td>
</tr>
<tr>
<td>Cyanazine</td>
<td>0.9 (0.3–2.5)</td>
<td>0.9 (0.4–1.9)</td>
<td>0.9 (0.5–1.6)</td>
<td>0.9 (0.3–1.5)</td>
<td>1.0 (0.5–1.3)</td>
</tr>
<tr>
<td>Metribuzin</td>
<td>0.9 (0.3–3.1)</td>
<td>0.9 (0.4–2.1)</td>
<td>1.0 (0.5–1.7)</td>
<td>0.9 (0.2–1.7)</td>
<td>1.0 (0.6–1.4)</td>
</tr>
<tr>
<td>EPTC</td>
<td>0.8 (0.2–3.4)</td>
<td>0.9 (0.4–2.2)</td>
<td>0.9 (0.5–1.7)</td>
<td>0.9 (0.3–1.9)</td>
<td>1.0 (0.6–1.5)</td>
</tr>
<tr>
<td>2,4,5-T</td>
<td>0.7 (0.1–3.2)</td>
<td>0.9 (0.3–2.0)</td>
<td>0.9 (0.5–1.7)</td>
<td>0.9 (0.3–1.7)</td>
<td>1.0 (0.6–1.3)</td>
</tr>
<tr>
<td>Atrazine</td>
<td>0.6 (0.2–1.4)</td>
<td>0.7 (0.3–1.3)</td>
<td>0.8 (0.5–1.3)</td>
<td>0.7 (0.1–1.3)</td>
<td>1.0 (0.4–1.2)</td>
</tr>
<tr>
<td>Metolachlor</td>
<td>0.5 (0.2–1.4)</td>
<td>0.6 (0.3–1.4)</td>
<td>0.8 (0.4–1.4)</td>
<td>0.8 (0.1–1.3)</td>
<td>1.0 (0.5–1.2)</td>
</tr>
<tr>
<td>Pendimethalin</td>
<td>0.5 (0.2–1.6)</td>
<td>0.7 (0.3–1.6)</td>
<td>0.8 (0.4–1.4)</td>
<td>0.8 (0.1–1.4)</td>
<td>1.0 (0.4–1.2)</td>
</tr>
<tr>
<td>Paraquat</td>
<td>0.3 (0.0–2.1)</td>
<td>0.7 (0.3–1.5)</td>
<td>0.8 (0.4–1.4)</td>
<td>0.8 (0.2–1.5)</td>
<td>1.0 (0.5–1.3)</td>
</tr>
</tbody>
</table>

*All models adjusted for state and age.

OR indicates ratio; CI stands for “confidence interval” for maximum likelihood and “credible interval” for Bayesian models; MLE, maximum likelihood estimate; 2,4,5-TP, 2,4,5-trichlorophenoxypropionic acid; 2,4,5-T, 2,4,5-trichlorophenoxyacetic acid; 2,4-D, 2,4-dichlorophenoxyacetic acid; EPTC, S-ethyl dipropylthiocarbamate.
current maximum likelihood techniques. The choice of model should be guided by careful thought regarding the likely magnitude of effects, as well as whether many effects of similar sizes may be seen.

REFERENCES
Innate Handedness and Disease-Specific Mortality in Women


Background: Left-handedness has been reported to be associated with reduced life expectancy, but the evidence is far from conclusive.

Methods: We studied the association between innate handedness and total mortality, as well as cause-specific mortality, in a cohort of 12,178 middle-aged Dutch women who were followed for almost 13 years. The relation between handedness and mortality was analyzed using Cox regression in a case-cohort approach, in which a random sample of 1500 women was used to represent person-years under observation for the entire cohort.

Results: During a median follow-up of 12.6 years, 252 women died. Hazard ratios comparing left-handed women with other women were 1.4 for all-cause mortality (95% confidence interval = 0.9–2.0), 1.7 for total cancer mortality (1.0–2.7), 2.0 for breast cancer mortality (0.8–4.6), 4.6 for colorectal cancer mortality (1.5–14.3), 1.3 mortality from diseases of the circulatory system (0.5–3.3), and 3.7 for cerebrovascular mortality (1.1–12.1), after adjusting for potential confounders (socioeconomic status, age, body mass index, and cigarette smoking status at study recruitment).

Conclusions: Left-handedness is associated with higher mortality in women.

(Epidemiology 2007;18: 208–212)

Approximately 10% of the population are left-handed.1 There is an increasing body of evidence in support of a genetic basis for hand preference,1,2 but many observations suggest environmental influences in early life as well.3,4 One of the prevailing explanations is that left-handedness stems from an adverse prenatal environment, such as excessive exposure to testosterone, which influences brain lateralization.3

In recent decades, there is a growing interest in early life origins of adult chronic diseases. Furthermore, childhood growth may modify the effect of birth weight on later life diseases risk5,6 and mortality.7 Hypotheses regarding fetal and infant origins of adult diseases8 may include mechanisms leading to both left-handedness and increased morbidity and mortality in later life.9 Just as an adverse environment in utero may lead to greater mortality from certain adult chronic diseases,10 left-handedness may similarly be associated with reduced survival. Indeed, left-handers are reported to be underrepresented in the older age groups, although such findings are still much debated.11,12

Several studies11,13–20 have attempted to explain the decreased number of left-handed people among elderly populations through a “modification” hypothesis, meaning cultural or social pressure against left-handedness, or an “elimination” hypothesis, meaning greater rates of mortality among the left-handed group.11 The modification hypothesis could not explain the reduced proportion of left-handedness in the elderly, as the estimated proportion of learned cases of left-handedness was much smaller than the estimated life-span shift toward dextrality.11 The elimination hypothesis has been studied to some extent, but the results are largely inconsistent.12 Two studies reported a relation between left-handedness and increased mortality,11,17 whereas 2 others found left-handedness to be associated with a survival benefit.18 However, many studies did not find any relation.13,16,19–21 Most of the studies above used cross-sectional designs, cohorts with short follow-up, or very selective groups of participants.

In this work, we explored the relationship between handedness and cause-specific mortality in a large prospective, population-based, cohort study of middle-aged women.

METHODS

In 1974, a population-based project was started in Utrecht, The Netherlands, and its surrounding municipalities, to study the early detection of breast cancer by mammographic screening.22 Within the project, there were 4 sub-birth cohorts. Women who participated in the third cohort (birth date 1932–1941) were asked by questionnaire about their handedness. Participation in this third cohort was approximately 40% from all invited women, with a total of 12,178 women recruited from 1982 to 1985.

These women filled out extensive questionnaires about reproductive history, demography, and lifestyle habits. They were asked about their innate hand preference: “Are you left or non left-handed by birth?” Of all women, 269 (2%) did not answer the question about handedness, and they were excluded from the analyses. Trained assistants took anthropo-
To assess the relation between handedness and mortality (overall and specific causes), we used weighted Cox regression analysis. The methods for such analysis are similar to standard Cox regression and previously have been described by Barlow et al.\textsuperscript{23} In the case–cohort design, the standard errors of risk estimates need to be corrected by a weighting scheme.\textsuperscript{23} We chose to use the weighting scheme proposed by Prentice\textsuperscript{25} because it was found to provide estimates that best resemble those from a full-cohort analysis.\textsuperscript{23} Follow-up time started from study inclusion (between 1982 and 1985) and ended at the date of death. Women who remained alive (or died of other causes when a specific cause was under investigation) during the observation period were censored at date of movement, date of death or at 1 January 1996, whatever occurred first.

Analyses were performed with SAS (version 8.2, SAS Institute Cary, NC) by use of a macro (available at http://lib.stat.cmu.edu/general/robphreg) that computes the weighted estimates together with a robust standard error, from which we calculated 95% confidence intervals (CIs). The proportionality of the hazards over time was evaluated with log-minus-log plots. Because cohort members were relatively healthy at recruitment, log-minus-log plots showed that mortality hazards were virtually identical between left- and non-left-handed women for the first 5 years after recruitment and started to deviate from each other afterward to reach parallelism. Therefore, to fully fulfill the proportional hazards assumption, we also analyzed the data excluding the first 5 years of follow-up. Because of limited numbers of events, this analysis was restricted only to overall mortality.

Univariate and multivariate models were run to consider potential confounders. Continuous variables were introduced into the models, and for categorical variables dummies were created. Hazard ratios (HRs) are reported with corresponding CIs.

RESULTS

At the end of follow-up, 90% of the random sample of 1455 women were still alive, 2.6% had died, 6.5% had migrated from the region, and 1.4% were lost to follow-up. A total of 17,567 person-years were accrued in the random sample, with a median time of follow-up of 151 months (12.6 years). Taking into account the sampling fraction, we extrapolated to 143,521 person years accrued in the total cohort, during which 240 women died (overall mortality rate: 1.7 per 1000 person-years).

Of the study population, 11.5% reported that they were left-handed “by birth.” Baseline characteristics according to handedness are presented in Table 1. Left- and non-left-handed women did not materially differ in age, BMI, SES, or smoking habits.

The left-handed group had a crude mortality rate for all causes of 2.3 (adjusted for age by direct standardization = 2.1) compared with 1.5 (adjusted for age = 1.6) per 1000 person-years in the non-left-handed group.

Table 2 shows that, after adjustment for age, SES, BMI, and cigarette smoking status, left-handed women had a 1.36 times higher risk of dying from all causes than non-left-
handed women. The adjusted HR for total mortality, after excluding the first 5 years of follow-up time, was 1.58 (95% CI = 1.03–2.42).

With regard to cancer mortality, left-handed women had a 1.7 times greater risk of dying from any type of cancer (CI = 1.0–2.7), a 4.6 times higher risk of dying from colorectal cancer (1.5–14), and a 2.0-fold higher risk of dying from breast cancer (0.83–4.6). Handedness was weakly associated with overall mortality from diseases of the circulatory system (1.3, 0.54–3.3), although left-handed women had a 3.7 times greater risk of dying from cerebrovascular diseases than non-left-handed women. Left-handedness was not associated with mortality attributable to causes other than the above-mentioned.

### DISCUSSION

In this cohort of middle-aged women, we have shown that left-handedness is associated with increased mortality, particularly from cancer and cerebrovascular diseases.

Certain features of our study need to be addressed. The participation rate in our cohort was 40%, and women were selected from participants in a breast cancer-screening project. Our population may be healthier compared with the general population, because women who voluntarily join screening programs are more likely to have healthier lifestyles and to be higher educated.

However, there is no reason to assume that the relationship between handedness and mortality would differ by participation.

### TABLE 1. Characteristics at Study Recruitment According to Innate Handedness

<table>
<thead>
<tr>
<th>Causes of Death</th>
<th>Random Sample</th>
<th>Whole Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Left-Handed (n = 168)</td>
<td>Non-Left-Handed (n = 1287)</td>
</tr>
<tr>
<td></td>
<td>Age at recruitment (yrs);</td>
<td></td>
</tr>
<tr>
<td></td>
<td>median (min, max)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>47.5 (41.6, 53.1)</td>
<td>46.9 (41.1, 53.1)</td>
</tr>
<tr>
<td></td>
<td>Age at the end of follow-up</td>
<td></td>
</tr>
<tr>
<td></td>
<td>years; median (min, max)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>58.8 (45.8, 64.0)</td>
<td>58.5 (42.8, 64.0)</td>
</tr>
<tr>
<td></td>
<td>BMI (kg/m²); mean ± SD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25.3 ± 4.0</td>
<td>24.8 ± 4.0</td>
</tr>
<tr>
<td></td>
<td>SES; no. (%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>98 (58)</td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>24 (14)</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>46 (27)</td>
</tr>
<tr>
<td></td>
<td>Cigarette smoking; no. (%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Never</td>
<td>88 (52)</td>
</tr>
<tr>
<td></td>
<td>Past (cigarettes/d) ≤ 20</td>
<td>16 (80)</td>
</tr>
<tr>
<td></td>
<td>&gt; 20</td>
<td>4 (20)</td>
</tr>
<tr>
<td></td>
<td>Current (cigarettes/d) ≤ 10</td>
<td>25 (42)</td>
</tr>
<tr>
<td></td>
<td>11–20</td>
<td>21 (35)</td>
</tr>
<tr>
<td></td>
<td>&gt; 20</td>
<td>14 (23)</td>
</tr>
</tbody>
</table>

### TABLE 2. Innate Handedness and Mortality

<table>
<thead>
<tr>
<th>Causes of Death</th>
<th>No. Deaths</th>
<th>Crude HR (95% CI)</th>
<th>Adjusted*HR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Left-Handed (16,985 PY)</td>
<td>Non-Left-Handed (126,509 PY)</td>
<td></td>
</tr>
<tr>
<td>All causes</td>
<td>37</td>
<td>203</td>
<td>1.43 (0.98–2.09)</td>
</tr>
<tr>
<td>All cancers</td>
<td>25</td>
<td>113</td>
<td>1.74 (1.10–2.74)</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>7</td>
<td>27</td>
<td>2.04 (0.88–4.69)</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>5</td>
<td>8</td>
<td>4.87 (1.57–15.05)</td>
</tr>
<tr>
<td>Other cancers</td>
<td>13</td>
<td>78</td>
<td>1.27 (0.69–1.73)</td>
</tr>
<tr>
<td>All diseases of the circulatory system</td>
<td>6</td>
<td>33</td>
<td>1.43 (0.59–3.45)</td>
</tr>
<tr>
<td>Cerebrovascular diseases</td>
<td>4</td>
<td>9</td>
<td>3.48 (1.06–11.39)</td>
</tr>
<tr>
<td>Other causes†</td>
<td>6</td>
<td>57</td>
<td>0.83 (0.35–1.93)</td>
</tr>
</tbody>
</table>

*Adjusted for SES, BMI, age at study recruitment, cigarette smoking status.
†Other causes: hepatitis, liver cirrhosis, metabolism disorder, meningitis, respiratory tract diseases, renal failure, injury.

PY indicates person-years extrapolated from the random sample.
A limitation of the present study is the small number of cases, which prohibited more detailed analyses on left-handedness and cause-specific mortality. Another limitation is that we could not study the effects of handedness in men, a group with a greater percentage of left-handedness.1

Measuring handedness using one question about writing hand or self-assessment may introduce misclassification.3 Because our focus originally was not on handedness, we asked only for innate hand preference (“at birth”). However, this method would presumably underestimate associations because this misclassification is likely to be at random (non differential). The proportion of innate left-handed people in our population is similar to previous studies.1

The data on mortality were obtained from active follow-up through regional municipalities and general practitioners, and we therefore believe it to be largely complete. The observed mortality rate in our cohort (2.3 per 1000 person-years) is lower than what would be expected from data from the general Dutch population (3.4 per 1000 person years).27 This rate probably reflects better average health among our participants.

We could think of only a limited number of possible confounders, in the sense of external factors biasing associations between handedness and mortality. Previously, it has been shown that age, BMI, SES, and cigarette smoking status are independent risk factors for mortality (overall and also for some specific causes of death)28,29 and some may also be related to hand preference.3 We adjusted for these factors (obtained at study recruitment), even though the left- and non-left-handed women showed only very minor differences with regard to these variables. Consequently, there were no material changes in the HRs after adjustment, leading us to conclude that in our data these factors seem to be neither confounders nor intermediate components of the causal pathway linking handedness to mortality. Finally, it is possible that our findings are due to chance.

There have been conflicting results in studies on the relationship between handedness and mortality. A study by Halpern and Coren17 found that left-handed people have a 9 years’ shorter life expectancy than their right-handed counterparts. Earlier studies11,12 also found a shorter life span in the left-handed group. These studies11,17 were widely criticized regarding their validity. More recent studies found no difference in mortality between the left- and non-left-handed subjects13,15,16,19,20 except for injury-related death.13,18 One found longer survival for the left-handed.18 All the aforementioned studies used current handedness, with various methods to measure handedness, in specific age and sex groups or in cohorts with shorter follow-up duration. The use of current handedness, especially in older generations could, as a result of societal pressure towards right-handedness, introduce misclassification of many innate left-handers as right-handers.

Left-handedness has been related to a higher occurrence of breast cancer in one case-control study30 and recently in our own cohort study31 and a lower occurrence of one type of brain tumor.32 To our knowledge, no previous study has shown a relationship between left-handedness and mortality because of cancer. Left-handedness was not related to cardiovascular disease mortality in previous studies.14,15,21

We could not examine the effect of left-handedness on injury-related mortality, because of a very small number of such events in our population. However, our study supports the hypothesis that fewer left-handed people among the elderly might be caused by elimination as the result of higher susceptibility to certain diseases or higher case-fatality among the left-handed, rather than an incapability to adapt to a right-handed world, leading to elimination by accident-related death.

If our observations about handedness and mortality are true, there could be several explanations. So far, the most plausible theory of handedness is the genetics theory from Annett et al10 and McManus et al,11 which still allows for a contribution from nongenetic factors. Nongenetic factors may cause some fraction of individuals to divert from their genetically designated handedness. The nongenetic factors may include exposure to an adverse environment, during fetal life or birth that leads to “atypical” laterality or pathologic handedness,3,4 disease susceptibility (inadequate immune system),3 and development of unhealthy lifestyles, such as smoking or alcoholism.33 However, we did not find clues that lifestyle differences between the handedness groups explained our findings. There is no evidence so far suggesting that genes involved in hand preference also act as an underlying factor for susceptibility for certain adult diseases.

In our study, left-handedness was related to a variety of major diseases with largely different causal mechanisms. Because breast cancer constitutes a majority of cancer deaths in our cohort, it dominates the relationship between left-handedness and cancer mortality. It has been suggested that breast cancer may be initiated in utero by high exposure to steroid hormones34 that may also cause left-handedness.3 For cardiovascular diseases, the link may be less straightforward. Nevertheless, as left-handedness is also more common in extremely low birth weight babies35 and low birth weight is related to cardiovascular mortality,36 the link might lay in the common intrauterine environment.37

In summary, the results of our study among 12,178 women followed for 13 years support the view that left-handedness is associated with higher mortality, especially as the result of cancer and cerebrovascular disease.

REFERENCES


**Occam’s Razor**

William of Occam (or Ockham) was a Franciscan monk in 14th Century England. His work as a logician and scholastic philosopher made him a well-known and provocative figure in academic circles. His name remains familiar today through Occam’s Razor, an expression of his philosophy that states “the simplest explanation is usually the best” (*Entia non sunt multiplicanda praeter necessitatem*). William died at the age of 64, felled in 1349 by the great wave of bubonic plague that killed nearly half the population of Europe.
Statin Use and the Risk of 10 Cancers

Patricia F. Coogan,* Lynn Rosenberg,,* and Brian L. Strom†‡

Background: Statins affect the proliferation, survival, and migration of cancer cells, and it is thought that they may have chemopreventive properties in humans. The purpose of the present study was to evaluate the association between statin use and various types of cancer in our hospital-based case–control surveillance study.

Methods: Data were collected from patients ages 40–79 years who were admitted to participating hospitals in 3 centers in Philadelphia, New York, and Baltimore from 1991 to 2005. Nurses administered questionnaires to obtain information on medication use and other factors. We compared patients who had any of 10 types of cancer (a total of 4913 patients) with controls admitted for noncancer diagnoses (3900 patients). The following cancers were examined individually: female breast (n = 1185), prostate (n = 1226), colorectal (n = 734), lung (n = 464), bladder (n = 240), leukemia (n = 254), pancreas (n = 220), kidney (n = 226), endometrial (n = 220), and non-Hodgkin lymphoma (n = 144). Logistic regression models were used to estimate odds ratios and 95% confidence intervals among regular statin users compared with never-users.

Results: Odds ratios were compatible with 1.0 for all cancer types. For the 4 largest cancer sites (breast, prostate, colorectum, and lung), odds ratios did not vary significantly by duration of statin use.

Conclusions: Statins are among the most commonly used medications, and durations of use are increasing. The present data do not support either positive or negative associations between statin use and the occurrence of 10 cancer types. Cancer incidence should continue to be monitored among statin users.

(Epidemiology 2007;18: 213–219)

Statin Use and the Risk of 10 Cancers

Statins were introduced to the U.S. market in 1987 and quickly eclipsed the older cholesterol-lowering medications. Atorvastatin (Lipitor; Pfizer, New York, NY) and simvastatin (Zocor; Merck, Whitehouse Station, NJ) were the 2 top-selling drugs in the United States in 2004.1 Statins are inhibitors of 3-hydroxy-3-methyl glutaryl coenzyme A (HMG-CoA) reductase, the enzyme required for conversion of HMG-CoA to mevalonic acid. Mevalonic acid is a precursor of cholesterol and of a variety of nonsterol isoprenoid derivatives that are important for various cellular functions, including cell proliferation, differentiation, and survival.2–4 Statins have been shown to arrest cell cycle progression,5,6 induce apoptosis,7 suppress angiogenesis, and alter the adhesion and migration of tumor cells7 in vitro and in vivo. Based on these findings, statins are thought to have chemopreventive potential in humans.8,9 Some observational data suggest that cholesterol levels themselves are inversely associated with the risk of cancer in humans,10–13 although low cholesterol may be a result of preclinical cancer.

A case–control study reported a 50% reduction in risk of colorectal cancer among statin users,14 and preliminary data from a Veterans Administration database found 50% reductions in risk for breast,15 prostate,16 and lung17 cancer. However, results from other studies do not support the hypothesis that statin use reduces the risk of cancer of the colon or breast or at other sites.18–25

We have previously evaluated the risk of breast and prostate cancer among statin users in our hospital-based case–control surveillance study.26 We found no associations of statin use with either prostate cancer or invasive breast cancer. Since the publication of those results, durations of use have increased. The purpose of the present study was to update the analyses for prostate cancer and invasive breast and to evaluate the association between statin use and other cancer types.

METHODS

Data were collected from patients admitted to participating hospitals in New York, Philadelphia, and Baltimore from 1991 to 2005. We did not include subjects interviewed from 1987 to 1990 because the prevalence of statin use was very low (<1%). The population base for the study comprised people living in zip codes within 50 miles of a participating hospital. Nurse-interviewers visited the hospitals on a rotating basis, from 1 to 3 times per week, depending on the size of the institution. On visit days, all patients meeting study criteria were identified through examination of admission lists and ward logs. Eligible patients were ages 18–79 years, were under the care of a physician participating in the study, did not have certain excluded diagnoses (eg, psychiatric), were able to complete the interview (eg, not deaf), and lived in an eligible zip code. Patients were approached for an interview if they did not have visitors or were not too ill.

Nurse-interviewers administered standard questionnaires to obtain information on demographic factors, medical
and reproductive history, and habits such as smoking and alcohol consumption. Lifetime history of medication use was elicited by asking about drugs used for 43 indications. We ascertained statin use by asking if any drugs had been used to lower blood cholesterol. We ascertained nonsteroidal anti-inflammatory drugs (NSAIDs) by asking about drugs used for pain, headache, arthritis, and other relevant indications. For each episode of drug use, the drug name and the duration, timing, and frequency of use were recorded. Subjects who could not specify what drug they took to lower blood cholesterol were not included in the analysis. Details of the diagnosis were abstracted from discharge summaries and pathology reports. From 1991 through 1997, an annual average of 90% of patients approached for an interview participated. Since 1998, an annual average of 82% of patients has participated. The study was approved by the institutional review boards of all participating institutions. All subjects signed informed consent forms.

We included as cases all patients ages 40 to 79 years with a primary cancer of a site for which there were at least 10 regular users of statins, whose cancer was first diagnosed within the year before the hospital admission at which the patient was interviewed, and who had no history of cancer or concurrent cancer other than nonmelanoma cancer of the skin. We analyzed the risk of cancer among regular statin users for 10 cancer types: invasive female breast (n = 1185), prostate (n = 1226), colorectal (n = 734), lung (n = 464), bladder (n = 240), leukemia (n = 254), pancreas (n = 220), kidney (n = 226), endometrial (n = 220), and non-Hodgkin lymphoma (n = 144).

Controls comprised 2081 women and 1819 men ages 40–79 years who were admitted to the hospital for diagnoses that we judged to be unrelated to statin use: orthopedic conditions (400 women and 422 men), trauma (with the exception of wrist and hip fractures; 399 women and 622 men), and other conditions (eg, cholelithiasis, hernia, acute infection; 1282 women and 775 men). Wrist and hip fractures were excluded because some studies have reported a reduced risk of fractures among statin users.27 Control patients had no history of cancer with the exception of nonmelanoma cancer of the skin. The age- and sex-adjusted prevalence of regular statin use (use at least 4 times a week for at least 3 continuous months) among controls has increased over time as follows: 1991–1993, 1.8%; 1994–1996, 4.7%; 1997–1999, 8.8%; 2000–2002, 13.4%; and 2003–2005, 16.3%. Among controls ages 50 years and older, the adjusted prevalence of use in 2003–2005 was 22.2%. Overall, the prevalence of regular statin use among subgroups of controls, adjusted for age (in 5-year categories), study center, and year of interview, was 4.8% for trauma and orthopedic conditions and 5.1% for other conditions. Among men, comparable figures were 5.0% and 6.4%; among women they were 4.6% and 4.3%.

Statistical Analysis

Regular statin use was defined as use at least 4 times per week for at least 3 continuous months, beginning at least 12 months before hospital admission. Regular use that began in the year before admission was considered to be etiologically irrelevant and was kept in a separate category. All other use was considered sporadic. The majority of the 41 cases and 20 control patients who reported sporadic statin use had used the drugs daily but for less than 3 months. NSAID use was defined as use of any drug in the following classes: salicylates (eg, aspirin), indoles (eg, indomethacin), propionic acids (eg, ibuprofen), fenamates (eg, mefenamic acid), pyrazolines (eg, phenylbutazone), and oxicams (eg, piroxicam). NSAID use categories were defined in the same way as for statin use; regular NSAID use was further divided into use that continued into the year before hospital admission (current use) and use that was stopped 1 year or more before hospital admission (discontinued use).

Unconditional logistic regression models were used to estimate odds ratios (ORs) and 95% confidence intervals (CIs) for the specific cancer types among regular statin users compared with never-users. Cancers that affected both men and women were compared with all controls, prostate cancer cases were compared with male controls, and breast and endometrial cancer cases were compared with female controls. For the endometrial cancer analyses, controls who had a hysterectomy were excluded. For all cancers, we present results from a basic model (model 1) adjusted for age in 5-year categories, interview year, study center, and sex (with the exception of sex-specific cancers). In addition we present ORs from a second model (model 2) adjusted for the variables listed above and for body mass index (kg/m²), alcohol consumption, race, years of education, pack-years of smoking, and NSAID use. For colorectal and prostate cancers, model 2 included family history of colorectal and prostate cancer, respectively. For breast and endometrial cancer, model 2 included use of conjugated estrogens or other female hormones (eg, progesterone, estradiol), use of oral contraceptives, menopausal status, parity, and age at menarche. For breast cancer we also included family history of breast cancer and religion. We classified prostate cancer by stage; this information was missing for 115 cases. We calculated ORs for categories of duration of use among regular users. To test for trend in the ORs across duration of use, we included duration of use as a continuous variable in the model that included regular users only.

RESULTS

Most cases and controls were white, and the majority of interviews were conducted before 1997 and in the largest study center, Philadelphia (Table 1). Cases tended to be older and more educated than control patients and were more likely to be ex-smokers. Table 2 shows the association of covariates with statin use among controls. Statin users were older than nonusers and were interviewed in later years of the study. Statin users were more likely to be white and to report regular NSAID use. Among men, statin users were less likely to be current smokers, and among women, statin users were less likely to be current drinkers, compared with nonusers. Statin use was not associated with family history of cancer of the colorectum, prostate, or breast. Among women, statin users were more likely to have used conjugated estrogens; other reproductive variables were not associated with statin use.
The fully adjusted odds ratios (model 2) were compatible with 1.0 for each of the 10 cancers considered. For the largest case groups of breast, prostate, colon, and lung, the ORs were 1.2 (95% CI = 0.8–1.8), 1.2 (0.9–1.7), 0.8 (0.5–1.2), and 0.7 (0.4–1.1), respectively (Table 3). ORs for cancers of the colon and rectum considered separately were also close to 1.0. Stage-specific odds ratios for prostate cancer were compatible with 1.0 (for stage I, 1.1 [CI = 0.6–2.0]; stage II, 1.3 [0.9–2.0]; stages III and IV, 1.3 [0.6–2.1]).

There were enough statin users among the colorectal, lung, breast, and prostate cancer cases for duration analyses (Table 4). We did not find any trends of increasing or decreasing ORs (model 2) across duration categories.

We calculated ORs for breast and prostate cancer among users of hydrophobic statins (ie, lovastatin, simvastatin, and fluvastatin). The fully adjusted OR for breast cancer among 43 case and 70 control users of the hydrophobic statins compared with nonusers of any statins was 1.02 (CI = 0.63–1.63), and 1.86 (0.93–3.74) among 26 case and 21 control users of the other statins. The fully adjusted odds ratio for prostate cancer among 98 case and 68 control users of hydrophobic statins was 1.29 (0.86–1.94), and 1.09 (0.62–1.92) among 55 case and 31 control users of other statins.

**DISCUSSION**

In these data, there was no pervasive evidence of associations between the regular use of statins and any type of cancer considered. There was a nonsignificant 30% decrease in the OR for lung cancer. However, the reduced risk was primarily in those using statins for less than 1 year, suggesting the association was not biologic. There was a nonsignificant 50% increase in the OR for breast cancer among women who had taken statins for 5 or more years, although with no other evidence of a trend.

Controls had diagnoses that we judged to be unassociated with statin use. We controlled for important potential confounding variables. However, control of confounding can not be considered complete in observational studies. Exposure classification was based on self-report, and recall bias is of concern. However, drugs taken to lower cholesterol were ascertained along with drugs taken for a long list of other indications, and the hypotheses concerning statins were unknown to both the interviewers and the participants. Because most subjects took statins daily and for long periods of time, their use is likely to have been reported satisfactorily, as suggested by 2 validation studies. In a case–control study of colorectal cancer,14 pharmacy records confirmed use for 276 (96.5%) of 286 self-reported statin users. In a case–control study of breast cancer among women ages 65–79 years,28 self-reported statin use at 3 time points before the index date was compared with pharmacy records. The sensitivity for use 6 months before the index date was 83% for cases and 93% for controls; comparable figures for 2 years before the index date were 75% and 86% and for 8 years before were 67% and 75%. Specificity was close to 100% for both cases and controls for the 3 time points.

Epidemiologic evidence on whether statins reduce cancer risk (as suggested by laboratory data) is mixed. A meta-analysis of 26 clinical trials found no association between statin use and overall cancer incidence or death.29 The ORs for specific cancer types (breast, prostate, gastrointestinal, colon, respiratory, and melanoma) did not vary from 1.0. The trials were meant to assess cardiovascular outcomes, however, and follow-up and power to assess cancer incidence was limited. In 4 studies that used data from pharmacy and other databases in either cohort or nested case-control analyses,18–21 the ORs for most cancer types were consistent with 1.0. The exceptions were a relative risk of 0.30 (95% CI = 0.11–0.81) for uterine cancer in a database from Quebec18 and an OR of 0.27 (95% CI = 0.8–0.95) for kidney cancer in a Dutch database.30 The database studies have had relatively short follow-up periods (ranging from a median of about 3 years18,19 to about 6 years20,21), but there were adequate numbers of case users for most cancer sites, especially for the larger sites of colorectal, prostate, and breast. In addition, the effect estimates were adjusted for a number of covariates. Researchers from the U.S. Veterans Administration have published abstracts on the association between statin use and cancer of the breast, prostate, and lung, using data on 1.4

---

**TABLE 1.** Characteristics of Cancer Cases and Controls, Case–Control Surveillance Study 1991–2005

<table>
<thead>
<tr>
<th></th>
<th>Cases (n = 4913)</th>
<th>Controls (n = 3900)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>47.9</td>
<td>46.6</td>
</tr>
<tr>
<td>Female</td>
<td>52.1</td>
<td>53.4</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>81.0</td>
<td>65.2</td>
</tr>
<tr>
<td>Nonwhite</td>
<td>19.0</td>
<td>34.8</td>
</tr>
<tr>
<td><strong>Study center</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Philadelphia</td>
<td>74.8</td>
<td>74.2</td>
</tr>
<tr>
<td>New York</td>
<td>14.1</td>
<td>10.6</td>
</tr>
<tr>
<td>Baltimore</td>
<td>11.1</td>
<td>15.2</td>
</tr>
<tr>
<td><strong>Interview year</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1991–1993</td>
<td>34.9</td>
<td>36.7</td>
</tr>
<tr>
<td>1994–1996</td>
<td>28.5</td>
<td>33.7</td>
</tr>
<tr>
<td>1997–2000</td>
<td>21.0</td>
<td>21.3</td>
</tr>
<tr>
<td>2001–2005</td>
<td>15.6</td>
<td>8.2</td>
</tr>
<tr>
<td><strong>Age (yrs)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40–49</td>
<td>21.5</td>
<td>40.9</td>
</tr>
<tr>
<td>50–59</td>
<td>32.8</td>
<td>31.7</td>
</tr>
<tr>
<td>60–69</td>
<td>37.3</td>
<td>22.2</td>
</tr>
<tr>
<td>70+</td>
<td>8.4</td>
<td>5.2</td>
</tr>
<tr>
<td><strong>Education (yrs)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;12</td>
<td>12.2</td>
<td>17.5</td>
</tr>
<tr>
<td>12</td>
<td>30.9</td>
<td>34.1</td>
</tr>
<tr>
<td>13–15</td>
<td>19.7</td>
<td>20.6</td>
</tr>
<tr>
<td>16+</td>
<td>37.1</td>
<td>27.6</td>
</tr>
<tr>
<td><strong>Smoking</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>36.8</td>
<td>36.7</td>
</tr>
<tr>
<td>Ex</td>
<td>40.2</td>
<td>33.1</td>
</tr>
<tr>
<td>Current</td>
<td>22.3</td>
<td>29.9</td>
</tr>
<tr>
<td></td>
<td>Men (n = 1819)</td>
<td>Women (n = 2081)</td>
</tr>
<tr>
<td>--------------------------</td>
<td>---------------</td>
<td>-----------------</td>
</tr>
<tr>
<td></td>
<td>No Statin Use</td>
<td>Regular Statin Use</td>
</tr>
<tr>
<td>Age (yrs)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40–49</td>
<td>39.1</td>
<td>20.2</td>
</tr>
<tr>
<td>50–59</td>
<td>33.3</td>
<td>30.3</td>
</tr>
<tr>
<td>60–69</td>
<td>23.2</td>
<td>33.3</td>
</tr>
<tr>
<td>70+</td>
<td>4.4</td>
<td>16.2</td>
</tr>
<tr>
<td>Interview year</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1991–1993</td>
<td>40.2</td>
<td>11.1</td>
</tr>
<tr>
<td>1994–1996</td>
<td>32.6</td>
<td>32.3</td>
</tr>
<tr>
<td>1997–2000</td>
<td>20.1</td>
<td>38.4</td>
</tr>
<tr>
<td>2001–2005</td>
<td>7.1</td>
<td>18.2</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>67.6</td>
<td>89.9</td>
</tr>
<tr>
<td>Nonwhite</td>
<td>32.4</td>
<td>10.1</td>
</tr>
<tr>
<td>NSAID use*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>39.4</td>
<td>25.3</td>
</tr>
<tr>
<td>Regular</td>
<td>17.0</td>
<td>44.4</td>
</tr>
<tr>
<td>Recent regular</td>
<td>4.2</td>
<td>2.0</td>
</tr>
<tr>
<td>Sporadic</td>
<td>39.1</td>
<td>28.3</td>
</tr>
<tr>
<td>Smoker</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>29.3</td>
<td>26.3</td>
</tr>
<tr>
<td>Ex-</td>
<td>37.9</td>
<td>54.5</td>
</tr>
<tr>
<td>Current</td>
<td>32.7</td>
<td>17.2</td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>21.9</td>
<td>30.3</td>
</tr>
<tr>
<td>Ex-</td>
<td>16.8</td>
<td>18.2</td>
</tr>
<tr>
<td>Current</td>
<td>61.4</td>
<td>51.5</td>
</tr>
<tr>
<td>Education (yrs)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;12</td>
<td>19.9</td>
<td>15.2</td>
</tr>
<tr>
<td>12</td>
<td>33.3</td>
<td>25.3</td>
</tr>
<tr>
<td>13–15</td>
<td>17.1</td>
<td>16.2</td>
</tr>
<tr>
<td>16+</td>
<td>29.6</td>
<td>43.4</td>
</tr>
<tr>
<td>Body mass index</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;21</td>
<td>5.5</td>
<td>2.0</td>
</tr>
<tr>
<td>21–24</td>
<td>26.3</td>
<td>20.2</td>
</tr>
<tr>
<td>25–29</td>
<td>43.3</td>
<td>47.5</td>
</tr>
<tr>
<td>30+</td>
<td>24.8</td>
<td>30.3</td>
</tr>
<tr>
<td>Family history of colorectal cancer</td>
<td>5.2</td>
<td>6.1</td>
</tr>
<tr>
<td>Family history of prostate cancer</td>
<td>3.2</td>
<td>5.1</td>
</tr>
<tr>
<td>Family history of breast cancer</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Ever use of conjugated estrogens</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Ever use of oral contraceptives</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>No. children</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>1–2</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>3–4</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>5+</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Age at first birth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>20–24</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>25–29</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>30+</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

*Regular use = at least 4 times/week for at least 3 months beginning ≥ 1 year before interview date. Recent regular use = regular use begun <1 year before interview date. Sporadic use = all other use.
### TABLE 3. Regular Statin Use Among Cancer Cases and Controls, Case–Control Surveillance Study, 1991–2005*

<table>
<thead>
<tr>
<th>Controls</th>
<th>No Statin Use (No.)†</th>
<th>Regular Statin Use (No.)</th>
<th>Model 1 Adjusted OR (95% CI)‡</th>
<th>Model 2 Adjusted OR (95% CI)§</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>3652</td>
<td>190</td>
<td>1.0 (0.8–1.6)</td>
<td>1.2 (0.8–1.8)</td>
</tr>
<tr>
<td>Females</td>
<td>1957</td>
<td>91</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Males</td>
<td>1695</td>
<td>99</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Cancer type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast</td>
<td>1101</td>
<td>69</td>
<td>1.1 (0.8–1.6)</td>
<td>1.2 (0.8–1.8)</td>
</tr>
<tr>
<td>Prostate</td>
<td>1027</td>
<td>153</td>
<td>1.4 (1.0–1.9)</td>
<td>1.2 (0.9–1.7)</td>
</tr>
<tr>
<td>Colorectum</td>
<td>691</td>
<td>35</td>
<td>0.8 (0.5–1.1)</td>
<td>0.8 (0.5–1.2)</td>
</tr>
<tr>
<td>Colon</td>
<td>446</td>
<td>21</td>
<td>0.7 (0.4–1.1)</td>
<td>0.7 (0.4–1.1)</td>
</tr>
<tr>
<td>Rectum</td>
<td>245</td>
<td>14</td>
<td>0.9 (0.5–1.6)</td>
<td>1.1 (0.6–2.0)</td>
</tr>
<tr>
<td>Lung</td>
<td>424</td>
<td>31</td>
<td>0.7 (0.5–1.1)</td>
<td>0.7 (0.4–1.1)</td>
</tr>
<tr>
<td>Bladder</td>
<td>216</td>
<td>20</td>
<td>1.2 (0.7–2.0)</td>
<td>1.3 (0.8–2.3)</td>
</tr>
<tr>
<td>Leukemia</td>
<td>238</td>
<td>15</td>
<td>1.1 (0.6–1.9)</td>
<td>1.1 (0.6–2.0)</td>
</tr>
<tr>
<td>Pancreas</td>
<td>208</td>
<td>10</td>
<td>0.7 (0.4–1.3)</td>
<td>0.7 (0.3–1.4)</td>
</tr>
<tr>
<td>Kidney</td>
<td>205</td>
<td>16</td>
<td>1.2 (0.7–2.0)</td>
<td>1.1 (0.6–1.9)</td>
</tr>
<tr>
<td>Endometrium</td>
<td>194</td>
<td>19</td>
<td>1.4 (0.8–2.5)</td>
<td>1.3 (0.7–2.4)</td>
</tr>
<tr>
<td>Non-Hodgkin</td>
<td>131</td>
<td>10</td>
<td>1.0 (0.5–2.1)</td>
<td>1.2 (0.6–2.4)</td>
</tr>
</tbody>
</table>

*Excluded 63 case and 37 controls who were recent statin users, 41 case and 20 controls who were sporadic statin users, and 3 cases and 1 control with unknown statin use.
†Reference category.
§Model 2: Adjusted for the variables listed above and body mass index (<21, 21–25, 25–30, 30+, unknown), alcohol use (never, ex-drinker, current drinker, unknown), race (white, all other races), years of education (<12, 12, 13–13, 16+, unknown), pack-years of smoking (never smoker, <10, 10–12, 24–44, 45+, unknown), and NSAID use (never, recent use, current use, discontinued use, sporadic use, unknown).
Colorectal cancer: also adjusted for family history of colorectal cancer (yes, no, unknown).
Prostate cancer: also adjusted for family history of prostate cancer (yes, no, unknown).
Breast and endometrial cancer: also adjusted for use of conjugated estrogens or other female hormones (never, <5 years, 5+ years, unknown), use of oral contraceptives (never, <5 years, 5+ years, unknown), menopausal status, parity (nulliparous, 1–2, 3–4, 5+, unknown), and age at menarche (<12, 12–14, 15+, unknown). Breast cancer OR also adjusted for family history of breast cancer (yes, no, unknown) and religion (Catholic, Jewish, Protestant, other, unknown).

### TABLE 4. Duration of Regular Statin Use Among Colorectal, Lung, Breast, and Prostate Cancer Cases and Control Patients*

<table>
<thead>
<tr>
<th>Controls</th>
<th>No Use†</th>
<th>≤1 Yr</th>
<th>&gt;1 to &lt;5 Yrs</th>
<th>5+ Yrs</th>
<th>P for Trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>3652</td>
<td>52</td>
<td>95</td>
<td>43</td>
<td>—</td>
</tr>
<tr>
<td>Female</td>
<td>1957</td>
<td>33</td>
<td>39</td>
<td>19</td>
<td>—</td>
</tr>
<tr>
<td>Male</td>
<td>1695</td>
<td>19</td>
<td>56</td>
<td>24</td>
<td>—</td>
</tr>
<tr>
<td>Cancer type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast</td>
<td>1101</td>
<td>17</td>
<td>32</td>
<td>20</td>
<td>1.5 (0.7–3.2)</td>
</tr>
<tr>
<td>Prostate</td>
<td>1027</td>
<td>28</td>
<td>68</td>
<td>54</td>
<td>1.2 (0.7–2.1)</td>
</tr>
<tr>
<td>Colorectum</td>
<td>691</td>
<td>8</td>
<td>13</td>
<td>12</td>
<td>1.1 (0.6–2.2)</td>
</tr>
<tr>
<td>Lung</td>
<td>424</td>
<td>4</td>
<td>16</td>
<td>10</td>
<td>0.9 (0.4–2.1)</td>
</tr>
</tbody>
</table>

*Cases missing date on duration of use: colorectum, 1; lung, 1; prostate, 3.
†Reference category.
‡Model 2: Adjusted for age (5-year categories), interview year (1991–1994, 1995–1999, 2000–2005), study center, sex (with the exception of breast and prostate), body mass index (<21, 21–25, 25–30, 30+, unknown), alcohol use (never, ex-drinker, current drinker, unknown), race (white, all other races), years of education (<12, 12, 13–13, 16+, unknown), pack-years of smoking (never smoker, <10, 10–12, 24–44, 45+, unknown), and NSAID use (never, recent use, current use, discontinued use, sporadic use, unknown).
Colorectal cancer: also adjusted for family history of colorectal cancer (yes, no, unknown).
Prostate cancer: also adjusted for family history of prostate cancer (yes, no, unknown).
Breast cancer: also adjusted for use of conjugated estrogens or other female hormones (never, <5 years, 5+ years, unknown), use of oral contraceptives (never, <5 years, 5+ years, unknown), menopausal status, parity (nulliparous, 1–2, 3–4, 5+, unknown), and age at menarche (<12, 12–14, 15+, unknown). Breast cancer OR also adjusted for family history of breast cancer (yes, no, unknown) and religion (Catholic, Jewish, Protestant, other, unknown).
million veterans enrolled in the South Central Veterans Administration Health Network. In these data, the ORs among statin users were 0.49 (95% CI = 0.38–0.62) for breast cancer15 0.46 (0.45–0.48) for prostate cancer,16 and 0.52 (0.49–0.55) for lung cancer.17 At this time few details are available with which to evaluate these results.

Four cohort and 2 case–control studies have evaluated statin use and its effect on breast or colon cancer risk. There was no association between breast cancer risk and statin use in the prospective Nurses Health Study cohort or in a case–control study of postmenopausal women.23 Another prospective study reported a relative risk for breast cancer of 0.28 (95% CI = 0.09–0.86) among statin users, but results were based on only 6 exposed cases.30 In the Women’s Health Initiative, overall use of statins was not associated with the risk of breast cancer (hazard ratio = 0.91 [95% CI = 0.80–1.05]), but there was a reduced risk among users of the hydrophobic statins (hazard ratio = 0.82 [0.70–0.97]).25 In at least one cell culture study, the hydrophobic statins induced apoptosis whereas the hydrophilic statin pravastatin did not.31 In our study, neither breast nor colon cancer risk was associated with use of the hydrophobic statins.

A case–control study of colorectal cancer conducted in Israel reported an odds ratio of 0.50 (95% CI = 0.40–0.63) among subjects who had used statins for at least 5 years compared with subjects who had used the drugs for less than 5 years.14 In the large Cancer Prevention Study II Nutrition cohort, however, there was no association between the risk of colorectal cancer and use of cholesterol-lowering drugs, the majority of which were statins.24

In several colorectal studies, cholesterol levels were inversely related to cancer incidence.10–12,32,33 In some studies, the strongest inverse associations were found during the first years of follow-up, suggesting that low cholesterol may be the result of preclinical cancer. Overall, evidence from the observational studies, as well as from dietary intervention studies and drug trials to lower cholesterol, does not make a strong case that lowering of cholesterol is associated with an increased cancer risk.13,34

In conclusion, the present data do not provide support for an association between statin use and the occurrence of most cancer types. In 2005, 24% of people ages 50 years and older reported using statins in the national Slone Survey of Medication Use.32 It is prudent public health policy to continue monitoring cancer incidence among users of such a commonly used drug class, particularly as durations of use are increasing.

REFERENCES


Cryptorchidism According to Maternal Gestational Smoking

Morten Søndergaard Jensen,* Gunnar Toft,* Ane Marie Thulstrup,* Jens Peter Bonde,* and Jørn Olsen†

Background: It has been suggested that maternal smoking during pregnancy is a risk factor for low sperm counts and testicular cancer in the offspring. Cryptorchidism is associated with both of these disorders and might share causal mechanisms.

Methods: We used prospective information on prenatal exposures and obstetric information on the birth of 5716 boys, collected from 1984 to 1987. During the 16–19 years of follow-up, 270 cases of cryptorchidism were diagnosed, and 185 of these boys underwent orchiopexy.

Results: Compared with nonsmokers, the adjusted risk ratio for being diagnosed with cryptorchidism was 1.1 (95% confidence interval = 0.8–1.6) if the mothers smoked 10–19 cigarettes/day and 2.3 (1.1–5.0) if they smoked ≥20 cigarettes/day. The risk ratios for orchiopexy were 1.4 (0.9–2.1) and 1.8 (0.6–5.0), respectively.

Conclusion: An excess risk of cryptorchidism was observed among sons of mothers who smoked 10 cigarettes or more per day during pregnancy. In recognition of the limited power of this study, the findings should be replicated in larger cohorts.

(EPIDEMIOLOGY 2007;18: 220–225)

Cryptorchidism is a common male congenital disorder, with an estimated prevalence at birth in Denmark as high as 9%.1 The known risk factors for cryptorchidism are low birth weight, short gestational age, and being small for gestational age.1–4 Cryptorchidism is associated with low sperm counts, infertility, and testicular malignancy.5–7

Recent findings suggest that maternal smoking during pregnancy is associated with lower sperm counts in the sons.8–10 Furthermore the finding of a strong ecologic association between maternal smoking and testicular cancer in offspring11 suggests a need to revisit Clemmesen’s hypothesis,12 that prenatal exposure to tobacco smoke is a cause of testicular cancer—despite the fact that earlier studies to evaluate this hypothesis have been negative.13–17

Considering the well-established association of maternal smoking during pregnancy with low birth weight,18–21 as well as the suggested associations with sperm counts and testicular cancer,8–11 it is reasonable to consider whether maternal smoking during pregnancy is also a risk factor for cryptorchidism. A recent study indicated that undescended testicular pathology is worsened by maternal smoking during pregnancy (decreased number of spermatogonia and gonocytes per tubule cross-section).22 Reports on prenatal smoking exposures and cryptorchidism to date show conflicting results, which is to be expected if the exposure effect is pregnancy time-dependent and also because the exposure data are crude and collected retrospectively.2,4,14,23–32 We present a follow-up study with prospectively collected data focusing on the relation between maternal tobacco smoking during pregnancy and cryptorchidism in the sons.

METHODS

In 1984–1987, the Danish birth cohort “Healthy Habits for Two” was established.33 Detailed information was collected on smoking habits and many other lifestyle exposures in pregnant women attending the midwife centers of Odense and Aalborg, Denmark. At their last scheduled regular antenatal care visit, around their 36th week of gestation, mothers filled in a comprehensive self-completed questionnaire and returned it by mail to the university Department of Social Medicine in Odense. Their medical records and questionnaires were given a unique study number and, after delivery, obstetric information was extracted from the medical records and linked to the questionnaire data by means of the study number. More than 80% of all pregnant women at the 36th week of gestation in the 2 defined geographical areas were enrolled—a total of 13,815 women. After excluding those lost to follow-up and those who gave birth to twins, data were available from the questionnaires and medical records for 11,144 women.

The women were asked about their smoking habits the year before pregnancy, smoking in early pregnancy, present smoking habits (36th week of gestation), type and brand of tobacco smoked, change in smoking habits during pregnancy, and the father’s smoking habits. The mothers’ smoking habits were classified according to the number of cigarettes smoked per day, and the cigarettes were classified according to the tobacco industries’ information on nicotine content per cig.
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Cryptorchidism Diagnosis</th>
<th></th>
<th>Orchiopexy</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. (%)</td>
<td>RR (95% CI)</td>
<td>No. (%)</td>
<td>RR (95% CI)</td>
</tr>
<tr>
<td>Mother</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking related to pregnancy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never*</td>
<td>124 (4.3)</td>
<td>1.0</td>
<td>79 (2.7)</td>
<td>1.0</td>
</tr>
<tr>
<td>Before pregnancy only</td>
<td>12 (4.6)</td>
<td>1.1 (0.6–2.0)</td>
<td>9 (3.4)</td>
<td>1.3 (0.6–2.6)</td>
</tr>
<tr>
<td>Stopped before week 36</td>
<td>15 (5.2)</td>
<td>1.2 (0.7–2.1)</td>
<td>12 (4.2)</td>
<td>1.6 (0.8–2.9)</td>
</tr>
<tr>
<td>Throughout pregnancy</td>
<td>116 (5.3)</td>
<td>1.2 (0.96–1.6)</td>
<td>82 (3.7)</td>
<td>1.4 (1.0–1.9)</td>
</tr>
<tr>
<td>Smoking in pregnancy (week 36)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No*</td>
<td>137 (4.3)</td>
<td>1.0</td>
<td>89 (2.8)</td>
<td>1.0</td>
</tr>
<tr>
<td>Yes</td>
<td>131 (5.2)</td>
<td>1.2 (0.96–1.6)</td>
<td>94 (3.8)</td>
<td>1.4 (1.0–1.8)</td>
</tr>
<tr>
<td>Number of cigarettes/d</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0*</td>
<td>153 (4.4)</td>
<td>1.0</td>
<td>102 (2.9)</td>
<td>1.0</td>
</tr>
<tr>
<td>1–9</td>
<td>51 (4.7)</td>
<td>1.1 (0.8–1.5)</td>
<td>36 (3.3)</td>
<td>1.1 (0.8–1.7)</td>
</tr>
<tr>
<td>10–19</td>
<td>56 (5.4)</td>
<td>1.2 (0.9–1.7)</td>
<td>42 (4.1)</td>
<td>1.4 (0.97–2.0)</td>
</tr>
<tr>
<td>≥20</td>
<td>9 (10.1)</td>
<td>2.4 (1.2–5.0)</td>
<td>4 (4.5)</td>
<td>1.6 (0.6–4.3)</td>
</tr>
<tr>
<td>Nicotine content of cigarettes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>9 (2.8)</td>
<td>0.5 (0.3–1.1)</td>
<td>5 (1.6)</td>
<td>0.4 (0.2–1.0)</td>
</tr>
<tr>
<td>Medium*</td>
<td>43 (5.2)</td>
<td>1.0</td>
<td>33 (4.0)</td>
<td>1.0</td>
</tr>
<tr>
<td>High</td>
<td>47 (6.0)</td>
<td>1.2 (0.8–1.8)</td>
<td>33 (4.2)</td>
<td>1.1 (0.7–1.8)</td>
</tr>
<tr>
<td>Age at delivery (yrs)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤19</td>
<td>15 (8.3)</td>
<td>2.0 (1.1–3.4)</td>
<td>14 (7.7)</td>
<td>2.9 (1.6–5.1)</td>
</tr>
<tr>
<td>20–29*</td>
<td>170 (4.4)</td>
<td>1.0</td>
<td>111 (2.9)</td>
<td>1.0</td>
</tr>
<tr>
<td>≥30</td>
<td>85 (5.2)</td>
<td>1.2 (0.9–1.6)</td>
<td>60 (3.7)</td>
<td>1.3 (0.9–1.8)</td>
</tr>
<tr>
<td>Infertility treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No*</td>
<td>234 (4.5)</td>
<td>1.0</td>
<td>159 (3.0)</td>
<td>1.0</td>
</tr>
<tr>
<td>Yes</td>
<td>34 (8.0)</td>
<td>1.9 (1.3–2.7)</td>
<td>25 (5.9)</td>
<td>2.0 (1.3–3.1)</td>
</tr>
<tr>
<td>Son</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birth weight (gm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤3000</td>
<td>63 (7.4)</td>
<td>2.0 (1.4–2.7)</td>
<td>44 (5.1)</td>
<td>1.9 (1.3–2.7)</td>
</tr>
<tr>
<td>3001–3500</td>
<td>94 (4.8)</td>
<td>1.3 (0.95–1.7)</td>
<td>59 (3.0)</td>
<td>1.1 (0.8–1.5)</td>
</tr>
<tr>
<td>&gt;3500*</td>
<td>113 (3.9)</td>
<td>1.0</td>
<td>82 (2.8)</td>
<td>1.0</td>
</tr>
<tr>
<td>Gestational age (wks)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤36</td>
<td>4 (3.7)</td>
<td>0.7 (0.3–2.0)</td>
<td>3 (2.8)</td>
<td>0.9 (0.3–2.7)</td>
</tr>
<tr>
<td>37–41*</td>
<td>194 (5.0)</td>
<td>1.0</td>
<td>127 (3.3)</td>
<td>1.0</td>
</tr>
<tr>
<td>&gt;41</td>
<td>69 (4.2)</td>
<td>0.8 (0.6–1.1)</td>
<td>54 (3.3)</td>
<td>1.0 (0.7–1.4)</td>
</tr>
<tr>
<td>Weight for gestational age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Light</td>
<td>14 (6.9)</td>
<td>1.5 (0.9–2.7)</td>
<td>9 (4.4)</td>
<td>1.4 (0.7–2.8)</td>
</tr>
<tr>
<td>Normal or heavy*</td>
<td>252 (4.6)</td>
<td>1.0</td>
<td>173 (3.2)</td>
<td>1.0</td>
</tr>
<tr>
<td>Placental weight (gm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤464</td>
<td>97 (6.8)</td>
<td>2.3 (1.6–3.4)</td>
<td>63 (4.4)</td>
<td>1.8 (1.2–2.7)</td>
</tr>
<tr>
<td>464–640</td>
<td>127 (4.5)</td>
<td>1.5 (1.0–2.1)</td>
<td>83 (2.9)</td>
<td>1.1 (0.8–1.7)</td>
</tr>
<tr>
<td>&gt;640*</td>
<td>43 (3.1)</td>
<td>1.0</td>
<td>36 (2.6)</td>
<td>1.0</td>
</tr>
<tr>
<td>Father</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking during pregnancy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No, never*</td>
<td>82 (3.9)</td>
<td>1.0</td>
<td>51 (2.4)</td>
<td>1.0</td>
</tr>
<tr>
<td>No, but previously</td>
<td>27 (3.9)</td>
<td>1.0 (0.7–1.6)</td>
<td>20 (2.9)</td>
<td>1.2 (0.7–2.1)</td>
</tr>
<tr>
<td>Yes</td>
<td>155 (5.5)</td>
<td>1.4 (1.1–1.9)</td>
<td>109 (3.9)</td>
<td>1.6 (1.2–2.3)</td>
</tr>
<tr>
<td>Number of cigarettes/d</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0*</td>
<td>109 (3.9)</td>
<td>1.0</td>
<td>71 (2.5)</td>
<td>1.0</td>
</tr>
<tr>
<td>1–10</td>
<td>45 (4.7)</td>
<td>1.2 (0.9–1.7)</td>
<td>31 (3.2)</td>
<td>1.3 (0.8–2.0)</td>
</tr>
<tr>
<td>11–20</td>
<td>64 (5.2)</td>
<td>1.3 (0.98–1.8)</td>
<td>46 (3.7)</td>
<td>1.5 (1.0–2.2)</td>
</tr>
<tr>
<td>≥20</td>
<td>20 (8.4)</td>
<td>2.7 (1.4–3.7)</td>
<td>14 (5.9)</td>
<td>2.4 (1.3–4.3)</td>
</tr>
<tr>
<td>Age at son’s birth (yrs)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤19</td>
<td>4 (11.8)</td>
<td>2.8 (0.98–8.1)</td>
<td>4 (11.8)</td>
<td>4.5 (1.6–13.2)</td>
</tr>
<tr>
<td>20–29*</td>
<td>128 (4.5)</td>
<td>1.0</td>
<td>81 (2.9)</td>
<td>1.0</td>
</tr>
<tr>
<td>30–39</td>
<td>114 (4.7)</td>
<td>1.0 (0.8–1.4)</td>
<td>82 (3.4)</td>
<td>1.2 (0.9–1.6)</td>
</tr>
<tr>
<td>≥40</td>
<td>15 (5.3)</td>
<td>1.2 (0.7–2.1)</td>
<td>10 (3.6)</td>
<td>1.3 (0.6–2.5)</td>
</tr>
</tbody>
</table>

*Reference category.
arettte, based on the reported brand names. In most cases, nicotine content was closely related to the tar content. The cigarettes’ nicotine content was grouped as follows: Low nicotine group (<1.5 mg/cigarette), medium nicotine group (1.5–1.9 mg/cigarette) and high nicotine group (≥2.0 mg/cigarette).

Case ascertainment was accomplished by means of the Danish National Patient Register and the unique 10-digit personal identifier in the Danish Civil Registration system. Of the 11,144 children, 5,716 were boys. All diagnoses or surgical procedures during hospital admissions of these boys (identified using the unique personal identifier) in the years 1984–2003 were extracted from the Danish National Patient Register.

In this study, we use 2 types of end points: boys diagnosed with cryptorchidism and boys who underwent orchiopexy (a subset of those with cryptorchidism). Orchiopexy is a specific marker for persistent cryptorchidism (thus requiring surgery), whereas the criteria for diagnosing cryptorchidism at birth are less well defined. Diagnosed cases were boys with the International Classification of Diseases, 8th edition (ICD-8) code 752.1 or any of the following ICD-10 codes: Q53, Q53.0, Q53.1, Q53.2, and Q53.9. (The ICD-9 codes were never officially used in Denmark). Orchiopexy cases were boys given the codes of KKHF00, KKHF01, or KKHF10 in the Nordic Classification of Surgical Procedures or the codes 5564 or 55640 in the Surgery and Treatment Classification of the Danish National Board of Health. The Danish Data Inspectorate approved the study.

We modeled the data through logistic regression using the Stata 9 statistical software package (StataCorp, College Station, TX). The reported risk ratios (RRs) were calculated as odds ratios (ORs) and presented with 95% confidence intervals (CIs). Potential confounders were identified in the literature. In addition to those presented in the Tables 1–4, several potential confounders were considered but not included in the analyses because they were not associated with either cryptorchidism (as judged by the risk estimates and confidence limits) or maternal smoking level (as judged by 2-way tabulation and χ² tests). These were son’s year of birth (cohort effect), cesarean delivery (no, acute, elective), number of white infarcts in the placenta, and birth complications (present or not present). Stratification of the association between maternal smoking during pregnancy and cryptorchidism is presented for nicotine content of the cigarettes and fathers’ daily smoking. The final models adjusted for variables that were either independent determinants of cryptorchidism or associated with maternal smoking levels. Because the intrauterine growth markers (birth weight, gestational age, and placental weight) may be intermediate variables in the association between maternal smoking and cryptorchidism 2 models were fitted. One model (model 1) adjusted for maternal and paternal age at delivery, time to index pregnancy, infertility treatment, parity, socioeconomic group and mothers’ alcohol intake. The other model (model 2) adjusted for the variables in model 1 plus the son’s birth weight, gestational age and placental weight. Furthermore, we did a crude analysis restricted to boys who had a birth weight >3500 g, were normal or heavy weight for gestational age, and who were born ≥37 weeks of gestation. There were 90 pairs of brothers among the 5716 boys (ie, same mother), and consequently these observations were not completely independent. We examined whether risk ratios changed after excluding them from the final models.

**RESULTS**

During the years of follow-up, 270 boys in the cohort were diagnosed with cryptorchidism (4.7%) and 185 of them underwent orchiopexy (3.2%). Cryptorchidism was diagnosed at birth for 18.5%, between birth and 18 months of age for 3.5%, and at ages older than 18 months for 78%. Table 1 shows the univariate distribution of end points according to mothers’ smoking habits and other selected risk factors. The following characteristics appeared to be associated with increased risk of cryptorchidism: the level of the mother’s smoking during pregnancy (including the nicotine content of the cigarettes), indicators of growth reduction in the son, infertility treatment, and the level of the father’s smoking during pregnancy. The fathers’ and mothers’ ages showed a u-shaped association with the risk of cryptorchidism, the risk being highest among the youngest. The data also suggest that the risk of having a boy with cryptorchidism is lower among those who quit smoking before the pregnancy.

**TABLE 2.** Crude and Adjusted Risk Ratios of Cryptorchidism Diagnosis or Orchiopexy, According to Mothers’ Daily Smoking

<table>
<thead>
<tr>
<th>Mothers’ Daily Smoking (Cigarettes/d)</th>
<th>Crude (n = 5683)</th>
<th>Model 1* (n = 5331)</th>
<th>Model 2† (n = 5226)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cryptorchidism RR (95% CI)</td>
<td>Orchiopexy RR (95% CI)</td>
<td>Cryptorchidism RR (95% CI)</td>
</tr>
<tr>
<td>0‡</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>1–9</td>
<td>1.1 (0.8–1.5)</td>
<td>1.1 (0.8–1.7)</td>
<td>1.0 (0.7–1.4)</td>
</tr>
<tr>
<td>10–19</td>
<td>1.2 (0.9–1.7)</td>
<td>1.4 (0.97–2.0)</td>
<td>1.2 (0.9–1.7)</td>
</tr>
<tr>
<td>≥20</td>
<td>2.4 (1.2–5.0)</td>
<td>1.6 (0.6–4.3)</td>
<td>2.2 (1.0–4.7)</td>
</tr>
<tr>
<td>Trend</td>
<td>1.2 (1.0–1.3)</td>
<td>1.2 (0.99–1.4)</td>
<td>1.1 (0.97–1.3)</td>
</tr>
</tbody>
</table>

*Model 1 adjusts for maternal and paternal age at delivery, time to index pregnancy, infertility treatment, parity, socioeconomic group, and mothers’ alcohol intake.
†Model 2 adjusts for model 1 variables plus birth weight, gestational age and placental weight (intrauterine growth markers).
‡Reference category.
§Relative risk increase per increase in exposure group of mothers’ daily smoking.
The crude risk ratios of cryptorchidism increased with maternal smoking during pregnancy (Table 2). Compared with nonsmokers, mothers reporting 1–9 cigarettes per day had a risk ratio of 1.1 (95% CI = 0.8–1.5); for 10–19 cigarettes per day, the RR was 1.2 (0.9–1.7), and for ≥20 cigarettes per day the RR was 2.4 (1.2–5.0). A positive trend of 1.2 (1.0–1.3) for each increase in smoking group was observed. The equivalent risk ratios for the orchiopexy cases were 1.1 (0.8–1.7), 1.4 (0.97–2.0), and 1.6 (0.6–4.3) for ≥20 cigarettes per day, with a positive trend of 1.2 (0.99–1.4). After adjustment (in both model 1 and model 2), there was no effect of smoking fewer than 10 cigarettes/day.

Table 3 presents the crude risk ratios of being diagnosed with cryptorchidism according to mothers' daily smoking stratified by nicotine content. As expected, very few of the heavy smokers (≥20 cigarettes/d) smoked cigarettes with medium or low nicotine content. There was a general trend for an increase in the number of cigarettes smoked or their nicotine (or tar) content to increase the risk of cryptorchidism. We noticed that, compared with nonsmokers, there was no excess risk if the mothers smoked 1–9 cigarettes a day with low or medium nicotine content. Similar results were seen for orchiopexy cases. Table 4 shows data stratified by the partner's smoking habits, and the result is compatible with an independent effect of each of the parents' smoking or an effect of environmental tobacco smoking.

We did a restricted analysis among term-born boys (n = 3396) who had a birth weight ≥3500 g, were normal or heavy weight for gestational age, and had a gestational duration of 37 weeks or longer. In this subgroup, we identified 141 cases of cryptorchidism. Compared with nonsmokers, the smoking of 1–9 cigarettes per day had a crude risk ratio of 0.9 (95% CI = 0.6–1.5), 10–19 cigarettes/day a RR of 1.4 (0.9–2.2), and ≥20 cigarettes per day a RR of 1.8 (0.5–5.9). There was a positive trend of 1.2 (0.94–1.4) for each increase in smoking group. Finally, we excluded the 90 pairs of brothers from the adjusted analysis, and the results were similar to those presented.

**DISCUSSION**

We observed an association between maternal tobacco smoking during pregnancy and cryptorchidism in the sons among both the diagnosed cases and the orchiopexy cases. After adjustment, the excess risk remained for the mothers who smoked ≥10 cigarettes per day. Furthermore, the data indicated that the nicotine content of the cigarettes may be of importance. In addition, an association between paternal smoking and cryptorchidism in the sons was observed. Since

<table>
<thead>
<tr>
<th>Nicotine Content</th>
<th>No. Cases</th>
<th>No. Boys</th>
<th>1–9 RR (95% CI)</th>
<th>10–19 RR (95% CI)</th>
<th>≥20 RR (95% CI)</th>
<th>Adjusted† RR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>9</td>
<td>321</td>
<td>0.3 (0.1–1.0)</td>
<td>1.3 (0.6–3.0)</td>
<td>—</td>
<td>1.0†</td>
</tr>
<tr>
<td>Medium</td>
<td>43</td>
<td>834</td>
<td>1.0 (0.6–1.6)</td>
<td>1.5 (1.0–2.3)</td>
<td>—</td>
<td>1.2 (0.9–1.5)</td>
</tr>
<tr>
<td>High</td>
<td>47</td>
<td>781</td>
<td>1.8 (1.2–2.7)</td>
<td>0.8 (0.5–1.5)</td>
<td>3.7 (1.4–9.8)</td>
<td>2.2 (1.0–4.5)</td>
</tr>
<tr>
<td>Adjusted†</td>
<td></td>
<td></td>
<td>1.0†</td>
<td>1.0 (0.7–1.5)</td>
<td>1.4 (0.5–3.6)</td>
<td></td>
</tr>
</tbody>
</table>

*Reference category for the primary analysis is nonsmokers (not shown).
†RR for nicotine content, adjusted for daily smoking.
‡RR for daily smoking, adjusted for nicotine content.
§Reference category.

**TABLE 4.** Crude Risk Ratios of Cryptorchidism Diagnoses, According to Mothers' Daily Smoking Stratified or Adjusted by Fathers' Daily Smoking and Vice Versa

<table>
<thead>
<tr>
<th>Father (Cigarettes/d)</th>
<th>No. Cases</th>
<th>No. Boys</th>
<th>0 RR (95% CI)</th>
<th>1–9 RR (95% CI)</th>
<th>10–19 RR (95% CI)</th>
<th>≥20 RR (95% CI)</th>
<th>Adjusted* RR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>109</td>
<td>2805</td>
<td>1.0†‡</td>
<td>1.0 (0.6–1.8)</td>
<td>1.3 (0.7–2.4)</td>
<td>1.3 (0.2–10)</td>
<td>1.0‡</td>
</tr>
<tr>
<td>1–10</td>
<td>45</td>
<td>958</td>
<td>1.1 (0.7–1.8)</td>
<td>1.5 (0.8–2.5)</td>
<td>1.1 (0.5–2.5)</td>
<td>4.6 (1.0–21)</td>
<td>1.2 (0.8–1.7)</td>
</tr>
<tr>
<td>11–20</td>
<td>64</td>
<td>1242</td>
<td>1.5 (0.9–2.3)</td>
<td>1.0 (0.5–1.9)</td>
<td>1.4 (0.9–2.3)</td>
<td>4.4 (1.5–13)</td>
<td>1.3 (0.9–1.8)</td>
</tr>
<tr>
<td>&gt;20</td>
<td>20</td>
<td>238</td>
<td>2.3 (1.0–5.1)</td>
<td>4.4 (1.7–12)</td>
<td>2.0 (0.9–4.5)</td>
<td>1.2 (0.2–9.1)</td>
<td>2.0 (1.2–3.4)</td>
</tr>
<tr>
<td>Adjusted†</td>
<td></td>
<td></td>
<td>1.0†</td>
<td>1.0 (0.7–1.5)</td>
<td>1.1 (0.8–1.6)</td>
<td>2.0 (0.9–4.2)</td>
<td></td>
</tr>
</tbody>
</table>

*RR for fathers' smoking level, adjusted for mothers smoking level.
†RR for mothers' smoking level, adjusted for fathers smoking level.
‡Reference category.
relative effect measures are biased downward by low diagnostic specificity, one would expect larger RRs for orchiopexy than for cryptorchidism, which was not seen. The reason could be that only persisting cases of cryptorchidism receive surgical treatment and that causal factors other than smoking contribute considerably to the persisting cases. Because the cohort was recruited late in gestation, we have few sons who were born preterm (1.9%) and none born very preterm among whom cryptorchidism is more common.

This study has strengths and limitations. The prospectively collected exposure information makes differential recall bias unlikely. Collecting smoking data around the 36th week of gestation may have introduced some nondifferential misclassification that could affect the power of the study, and bias the results towards null values. The case ascertainment was registry-based, and it is well known that not all cases are diagnosed. If maternal tobacco smoking prompts a clinician to do a more thorough search for cryptorchidism, information bias could lead to high risk estimates. We find this unlikely, because the prevalence of congenital malformations among smoking-exposed newborns is similar to the prevalence among babies who are not exposed to smoking.\(^3\) Furthermore, the risk estimates were replicated when we restricted the analysis to boys born at term, which makes information bias dependent on prematurity very unlikely. This mechanism of bias would be strongest while the boys are very young; we followed all boys up to the age of 16 to 19 years.

The sensitivity and specificity of the registration of cryptorchidism cases has, to our knowledge, not been directly validated. The cryptorchidism prevalence in this cohort was somewhat greater than the 2.0% reported in another Danish register study covering the same time period (1983–1992),\(^3\) probably because their follow-up period was only 9 years. If we truncated our data at 9 years of follow-up, the prevalence of the diagnosis was 3.3% (2.2% for orchiopexy).

The relatively strong association with heavy smoking (≥20 cigarettes per day), in addition to the possible effect of the cigarettes’ nicotine content, indicate that the association between maternal smoking during pregnancy and cryptorchidism may be causal; the reduced risk for those who quit smoking before getting pregnant is consistent with this interpretation. The fact that the association was relatively robust despite adjustments also speaks in favor of causality, but the association between the fathers’ smoking and cryptorchidism does not. This association could be a simple consequence of the strong correlation between maternal and paternal smoking, or perhaps due to a large underreporting by mothers of their own smoking habits when their partners smoked. The association may also suggest an effect of passive smoking. To pursue this it would be useful to conduct a study with questionnaire information on both parents’ smoking habits and an internal measure of the mothers’ smoking (serum cotinine).

Previous studies on the association between maternal smoking and cryptorchidism have mostly been of the case-control type; with no information on the level of smoking.\(^2,4,14,25-32\) They have reported odds ratios for cryptorchidism between 1.0 and 1.7. If our findings are true, the prevalence of women smoking 10 cigarettes or more per day will determine the excess risk in a group of smokers, and only few populations have smoking levels that can match what was seen for Danish women in the 1980s. Prenatal tobacco exposure have also been reported as a risk factor for low sperm counts.\(^8-10\) Interestingly, the studies that reported maternal smoking dose generally found an adverse effect when the mothers smoked 10 cigarettes per day or more.\(^8,10\)

The observed robustness with adjustment for birth weight has also been described in other studies.\(^10,25\) Together, this evidence suggests that tobacco components may cause specific damage to the male embryonic gonads. This is supported by a recent observation that maternal gestational smoking worsens the testicular pathology in cryptorchidism.\(^22\) These studies add to the suspicion that prenatal exposure to tobacco smoking may be one of the causes of testicular cancer. If this cancer is causally related to gestational-time-specific prenatal exposure to tobacco smoking, with a second carcinogenic hit later in life, the prenatal exposure association would be attenuated and perhaps even eliminated in case of negative confounding for smoking related to the second hit. This might also explain why the studies on individuals show ambiguous results while ecologic studies have given more suggestive associations.\(^11,13-17\)

In conclusion, we observed an excess risk of cryptorchidism for mothers who smoked at least 10 cigarettes per day during pregnancy. In recognition of the limited power of this study, the findings should be replicated in larger cohorts.

ACKNOWLEDGMENTS

The authors thank everyone who participated in establishment of the “Healthy Habits for Two” cohort. Special thanks go to Bodil Hammer Bech and Gitte Nielsen at the Danish Epidemiology Science Centre.

REFERENCES


Maternal Smoking and Environmental Tobacco Smoke Exposure and the Risk of Orofacial Clefts

Margaret A. Honein,* Sonja A. Rasmussen,* Jennita Reehuis,* Paul A. Romitti,†
Edward J. Lammer,‡ Lixian Sun,† and Adolfo Correa*

Background: Smoking during pregnancy has been associated with orofacial clefts in numerous studies. However, most previous studies have not been able to assess the relation between maternal smoking and specific phenotypes (eg, bilateral clefts).

Methods: We examined the association between periconceptional maternal smoking, environmental tobacco smoke (ETS) exposure, and cleft lip with or without cleft palate (CLP) (n = 933) and cleft palate only (CPO) (n = 528) compared with infants with no major birth defects (n = 3390). Infants were born between 1 October 1997 and 31 December 2001, and exposures were ascertained from maternal telephone interviews for the National Birth Defects Prevention Study. We excluded infants who had a first-degree relative with an orofacial cleft. Effect estimates were adjusted for folic acid use, study site, prepregnancy obesity, alcohol use, gravidity, and maternal age, education, and race/ethnicity.

Results: Periconceptional smoking was associated with CLP (odds ratio = 1.3; 95% confidence interval = 1.0–1.6), and more strongly associated with bilateral CLP (1.7; 1.2–2.6), with a weaker association observed for CPO. Heavy maternal smoking (25 cigarettes/day) was associated with CLP (1.8; 1.0–3.2), bilateral CLP (4.2; 1.7–10.3), and CPO with Pierre Robin sequence (2.5; 0.9–7.0). ETS exposure was not associated with CLP or CPO.

Conclusions: This study confirmed the modest association between smoking and orofacial clefts that has been consistently reported, and identified specific phenotypes most strongly affected.

(O Epidemiology 2007;18: 226–233)

METHODS

Orofacial clefts are among the most common major birth defects in the United States, with approximately 1 in 870 live births affected by cleft lip with or without cleft palate (CLP), and 1 in 1500 births affected by cleft palate only (CPO).1 Previous studies have identified a number of factors associated with orofacial clefts, including family history; genetic factors; birth order; occupational exposures; anticonvulsants; multivitamin intake, alcohol, and smoking.11–16

The association between maternal smoking and orofacial clefts has been assessed in many studies, and a meta-analysis of these studies suggests a modest positive association for both CLP and CPO.17 However, previous studies have had a number of limitations, including inability to examine a dose–response effect of smoking, inadequate sample size to control for confounders, lack of careful case review and classification of cases, insufficient sample size or clinical detail to examine bilateral and unilateral CLP separately, and insufficient numbers to examine CPO with and without Pierre Robin sequence. The ability to separate the cases into more homogeneous subgroups is critical to understanding the possible mechanisms of a smoking effect. In addition, few previous studies have been able to assess the role of environmental tobacco smoke (ETS) exposure in the etiology of orofacial clefts. The objective of this analysis is to assess the association between maternal smoking, maternal exposure to ETS, and the occurrence of orofacial clefts in a large population-based case-control study of major birth defects in the United States.

From the *National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, Georgia; †University of Iowa, Iowa City, Iowa; and ‡Children’s Hospital Oakland Research Institute, Oakland, California.

Supported by the Centers for Disease Control and Prevention. Supported, in part, by the Cigarette and Tobacco Surplus Fund of the California Tobacco-Related Diseases Research Program, University of California, grant #13RT-0109.

The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

Correspondence: Margaret A. Honein, Centers for Disease Control and Prevention, 1600 Clifton Road, NE, Mailstop E-86, Atlanta, GA 30333.

E-mail: Mhonein@cdc.gov

Copyright © 2007 by Lippincott Williams & Wilkins
ISSN: 1044-3983/07/1802-0226
DOI: 10.1097/01.ede.0000254430.61294.c0
chromosome abnormality. All sites ascertain case infants diagnosed within 12 months after delivery, and some sites ascertain case infants diagnosed beyond one year of age. Control infants are a random sample of live births (with no major birth defects) from the same population as the case infants. The source population is defined by maternal residence at delivery. Control infants are selected proportionate to the number of births occurring in a given month; some sites select control infants from birth certificate files and some from delivery logs of birth hospitals. Trained interviewers conduct telephone interviews with each mother to collect information on the following exposures during the 3 months prior to conception and during pregnancy: maternal illnesses and medication; pregnancy history; nutrition; caffeine, tobacco, alcohol, and illicit drugs; occupation and environment; and demographics. Seventy-six percent of eligible orofacial cleft case mothers and 69% of eligible control mothers participated in the interview.

In addition to the routine case review by clinical geneticists at each site, all infants with a diagnosis of an orofacial cleft were reviewed by one clinical geneticist (S.A.R.) to ensure that they met the inclusion criteria for this analysis. Infants with an orofacial cleft secondary to another defect (eg, holoprosencephaly or amniotic band sequence) were excluded. All infants were classified as either having an isolated cleft (no additional major defect) or multiple defects (major unrelated defects in at least 2 different organ systems), because these groups likely have different causes, and infants with multiple defects are more likely to have an unrecognized genetic condition than infants with isolated defects.19

Orofacial clefts were analyzed in 2 major categories (CLP and CPO) because of the presumed etiologic and pathogenetic distinctions between these malformations. Further subgroups for analysis included CPO, with and without the diagnostic code for Pierre Robin sequence in their abstracted medical record, isolated and multiple, and males and females. Subanalyses of CLP included unilateral and bilateral, isolated and multiple, males and females and isolated/bilateral and isolated/unilateral. These further subanalyses enhance our capacity to explore possible etiologic mechanisms. For example, CPO in infants with Pierre Robin sequence might be a result of a pathogenetic mechanism involving the mandible, in contrast to cases having primary clefts of the palate.

All mothers were asked whether they had ever smoked cigarettes and, if so, whether they had smoked cigarettes any time from 3 months before pregnancy until their infant’s birth. Mothers who reported any smoking in this time period were asked specifically whether they smoked during each of the following 8 time periods: the third, second, or first months before pregnancy; the first, second, and third months of pregnancy; and the second and third trimesters of pregnancy. Mothers were also asked to report the number of cigarettes they smoked per day during each of these time periods.

Mothers reported their exposure to environmental tobacco smoke at home or work during the same time periods. Mothers were asked whether anyone in their household smoked cigarettes in their home during their pregnancy or the 3 preceding months and about each specific time period. Next, mothers were asked whether anyone smoked cigarettes near them at a workplace or school during the same time periods.

Infants were classified as exposed to periconceptional maternal smoking if their mother reported smoking at any time in the month before or the first 3 months of pregnancy. The referent or unexposed comparison group for our assessment of maternal smoking was infants whose mother reported no smoking and no exposure to ETS at home or work during the month before pregnancy or the first 3 months of pregnancy. Among women reporting periconceptional smoking, the level of smoking was classified as heavy (25+ cigarettes per day), medium (15–24 cigarettes per day), or light (1–14 cigarettes per day).

Our assessment of ETS was limited to nonsmoking mothers, defined as mothers who reported no active smoking during the month before or the first 3 months of pregnancy. Infants were classified as exposed to ETS if their mother reported any exposure to household or workplace/school tobacco smoke in the month before pregnancy or the first 3 months of pregnancy. Infants whose mothers reported no ETS exposure during the same period were the referent.

We included infants born on or after 1 October 1997 up through infants with an estimated date of delivery on or before 31 December 2001. Analyses were limited to mothers with completed interviews; 13 case mothers and 35 control mothers with partial interviews were excluded. All main effects were assessed by crude and stratified analyses using SAS version 9.0 (SAS Institute Inc., Cary, NC). To control for potential confounding, an unconditional logistic regression model was fit to the data. We excluded infants having a first-degree relative with an orofacial cleft (14 controls, 52 CLP, 28 CPO), according to maternal report. All effect estimates were adjusted for the following variables, selected a priori based on biologic plausibility and prior studies,14,15 self-reported maternal race/ethnicity (non-Hispanic white, non-Hispanic black, Hispanic, other), study center (8 sites), periconceptional folic acid use (any use in the month before pregnancy or the first month of pregnancy versus no use), maternal age (up to 30 years, 31 years or older), infant sex, prepregnancy obesity (body mass index $\geq 30$ kg/m$^2$, <30 kg/m$^2$), first-trimester alcohol use (any use in the month before or the first 3 months of pregnancy versus no use), maternal education (0–12 years, >12 years), and gravidity (primigravid, multigravid).

RESULTS

Mothers of 933 infants with nonsyndromic CLP, 528 infants with nonsyndromic CPO, and 3390 infants with no major birth defects (controls) completed the maternal telephone interview. The majority of case infants had no other defects: 88% of infants with CLP and 80% of infants with CPO. Among infants with CLP, 67% had unilateral CLP, 22% had bilateral CLP, 1% had central CLP, and 10% had CPO with no information on laterality. One quarter of infants with CPO had a diagnosis of Pierre Robin sequence (Table 1).

The majority of infants with CPO were male (66%), and just over half of infants with CPO were female (53%). Hispanic ethnicity was associated with CLP and, consequently, the 2 study sites with the highest representation of Hispanic mothers

© 2007 Lippincott Williams & Wilkins

© 2007 Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.
associated with CLP (adjusted odds ratio 1.8; 95% confidence interval [CI] = 1.0–2.9) than were women who did not smoke during this period. The association with heavy smoking was strongest for infants with CPO without Pierre Robin sequence (1.3; 0.9–2.0), whereas a much weaker association was observed among infants with CLP with Pierre Robin sequence (1.2; 0.9–7.0), whereas a much weaker association was observed among infants with CLP without Pierre Robin sequence. The effects of maternal smoking were similar for male and female infants (Table 3).

Effect estimates were greater for infants exposed to the heaviest levels of maternal smoking (Table 4). Mothers who smoked heavily in the periconceptional period were about twice as likely to have an infant with any orofacial cleft (1.8; 95% CI 1.1–2.9) than were women who did not smoke during this period. The association with heavy smoking was strongest for infants with bilateral CLP (4.2; 95% CI 1.7–10) and CLP with multiple defects (3.1; 1.0–10). CPO with Pierre Robin sequence was associated with heavy maternal smoking (2.5; 95% CI 0.9–7.0), whereas a much weaker association was observed among infants with CPO without Pierre Robin sequence (1.4; 0.6–3.2). Among mothers with no folic acid use periconceptionally, the association with smoking was stronger (1.4; 1.1–1.8) than among mothers who took folic acid (1.2; 0.9–1.5).

Among nonsmoking mothers, exposure to any ETS at home or work in the periconceptional period was not associated overall with either CLP or CPO (Table 5). There was a modest association with ETS among infants with CPO with multiple defects (1.7; 1.0–3.0), and a weak association among female infants with CPO (1.3; 0.9–2.0). No dose information was available to quantify the level of ETS exposure.

### DISCUSSION

In this study, periconceptional maternal smoking was associated with orofacial clefts and in particular CLP. The effect was strongest for bilateral CLP and isolated CLP. Among infants of the heaviest smoking mothers, bilateral CLP was 4 times more likely and CPO with Pierre Robin sequence was over twice as likely as among nonsmoking mothers. However, only 2.2% of case-mothers and 1.3% of control-mothers reported this heaviest level of smoke exposure. Overall, the heaviest smokers were about twice as likely to have an infant with an orofacial cleft. There was no overall impact of ETS exposure on the occurrence of orofacial clefts. However, we were not able to assess ETS exposure by dose of the exposure, and any true effect might have been masked by mixing light and heavy ETS exposures. Our findings are consistent with a recent meta-analysis of smoking and orofacial clefts that included 24 studies published between 1974 and 2001.17 For CLP and CPO, the meta-analysis yielded effect estimates essentially the same as our study. Assuming the relation between maternal smoking and orofacial clefts is causal, 4% of all orofacial clefts and 12% of bilateral CLP can be attributed to periconceptional maternal smoking.

Heavy maternal smoking was associated with CPO in the presence of the Pierre Robin sequence but not with other cases of CPO. Palatal clefts arise from failure of the palatal shelves to fuse, which might be due to several mechanisms, for example, a failure of elevation of the palatal shelves or micrognathia and resultant posterior placement of the tongue impairing closure of the shelves (Pierre Robin sequence). Evaluation of the effect of smoking in the presence and absence of Pierre Robin may help to identify the mechanism responsible for the association. Heavy maternal smoking was also associated with CLP (OR 1.8) and most strongly asso-

---


<table>
<thead>
<tr>
<th>Defect</th>
<th>Total</th>
<th>Isolated Defect</th>
<th>Multiple Defects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>(%)</td>
<td>No.  (%)</td>
</tr>
<tr>
<td>CLP</td>
<td>933</td>
<td>819</td>
<td>114</td>
</tr>
<tr>
<td>Bilateral</td>
<td>204</td>
<td>22</td>
<td>21</td>
</tr>
<tr>
<td>Unilateral</td>
<td>625</td>
<td>67</td>
<td>69</td>
</tr>
<tr>
<td>Central</td>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Not otherwise specified</td>
<td>97</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>CPO</td>
<td>528</td>
<td>422</td>
<td>106</td>
</tr>
<tr>
<td>With Pierre Robin sequence</td>
<td>133</td>
<td>25</td>
<td>24</td>
</tr>
<tr>
<td>Without Pierre Robin sequence</td>
<td>395</td>
<td>75</td>
<td>76</td>
</tr>
</tbody>
</table>

CLP indicates cleft lip with or without cleft palate; CPO, cleft palate alone.
### TABLE 2. Descriptive Information About Infants With CLP, CPO, and all Controls

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Controls (n = 3390)</th>
<th>CLP (n = 933)</th>
<th>CPO (n = 528)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number (%)*</td>
<td>Unadjusted OR (95% CI)</td>
<td>Number (%)*</td>
</tr>
<tr>
<td><strong>Infant sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female†</td>
<td>1712 (51)</td>
<td>2.0 (1.8–2.4)</td>
<td>282 (53)</td>
</tr>
<tr>
<td>Male</td>
<td>1674 (49)</td>
<td></td>
<td>245 (46)</td>
</tr>
<tr>
<td><strong>Maternal race/ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic white†</td>
<td>2072 (61)</td>
<td>3.62 (69)</td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic black</td>
<td>392 (12)</td>
<td>0.5 (0.3–0.7)</td>
<td>33 (6)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>756 (22)</td>
<td>1.2 (1.0–1.4)</td>
<td>96 (18)</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>88 (3)</td>
<td>1.1 (0.8–1.4)</td>
<td>20 (4)</td>
</tr>
<tr>
<td>Native American/Alaskan Native</td>
<td>14 (0)</td>
<td>2.1 (0.8–5.3)</td>
<td>3 (1)</td>
</tr>
<tr>
<td>Other</td>
<td>38 (1)</td>
<td>1.2 (0.6–2.4)</td>
<td>5 (1)</td>
</tr>
<tr>
<td><strong>Study site</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arkansas</td>
<td>401 (12)</td>
<td>0.9 (0.7–1.3)</td>
<td>54 (10)</td>
</tr>
<tr>
<td>California</td>
<td>468 (14)</td>
<td>1.4 (1.0–1.9)</td>
<td>64 (12)</td>
</tr>
<tr>
<td>Georgia†</td>
<td>377 (11)</td>
<td>73 (14)</td>
<td></td>
</tr>
<tr>
<td>Iowa</td>
<td>402 (12)</td>
<td>1.1 (0.8–1.5)</td>
<td>50 (9)</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>425 (13)</td>
<td>1.1 (0.8–1.5)</td>
<td>94 (18)</td>
</tr>
<tr>
<td>New Jersey</td>
<td>497 (15)</td>
<td>0.8 (0.6–1.1)</td>
<td>68 (13)</td>
</tr>
<tr>
<td>New York</td>
<td>375 (11)</td>
<td>1.0 (0.7–1.4)</td>
<td>62 (12)</td>
</tr>
<tr>
<td>Texas</td>
<td>445 (13)</td>
<td>1.4 (1.0–1.8)</td>
<td>63 (12)</td>
</tr>
<tr>
<td><strong>Gestational age</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very preterm (&lt;32 wk)</td>
<td>33 (1)</td>
<td>2.7 (1.5–4.8)</td>
<td>20 (4)</td>
</tr>
<tr>
<td>Preterm (32–36 wk)</td>
<td>259 (8)</td>
<td>1.7 (1.4–2.2)</td>
<td>80 (15)</td>
</tr>
<tr>
<td>Term (37–45 wk)†</td>
<td>3095 (91)</td>
<td>427 (81)</td>
<td></td>
</tr>
<tr>
<td><strong>Maternal age at delivery (yrs)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;18</td>
<td>149 (4)</td>
<td>1.0 (0.7–1.5)</td>
<td>18 (3)</td>
</tr>
<tr>
<td>18–19</td>
<td>246 (7)</td>
<td>1.5 (1.1–2.0)</td>
<td>37 (7)</td>
</tr>
<tr>
<td>20–24</td>
<td>728 (21)</td>
<td>1.3 (1.1–1.6)</td>
<td>112 (21)</td>
</tr>
<tr>
<td>25–29†</td>
<td>878 (26)</td>
<td>1.2 (1.1–1.6)</td>
<td>127 (24)</td>
</tr>
<tr>
<td>30–34</td>
<td>918 (27)</td>
<td>0.8 (0.7–1.0)</td>
<td>141 (27)</td>
</tr>
<tr>
<td>35–39</td>
<td>396 (12)</td>
<td>1.1 (0.8–1.4)</td>
<td>78 (15)</td>
</tr>
<tr>
<td>40–44</td>
<td>67 (2)</td>
<td>1.1 (0.6–1.9)</td>
<td>14 (3)</td>
</tr>
<tr>
<td>45–49</td>
<td>7 (0)</td>
<td>1.7 (0.3–7.2)</td>
<td>1 (0)</td>
</tr>
<tr>
<td><strong>Maternal education (yrs)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–8</td>
<td>150 (4)</td>
<td>1.2 (0.8–1.7)</td>
<td>32 (6)</td>
</tr>
<tr>
<td>9–11</td>
<td>402 (12)</td>
<td>1.2 (0.9–1.5)</td>
<td>62 (12)</td>
</tr>
<tr>
<td>12†</td>
<td>877 (26)</td>
<td>1.2 (0.9–1.5)</td>
<td>134 (25)</td>
</tr>
<tr>
<td>13–15</td>
<td>944 (28)</td>
<td>0.8 (0.7–1.0)</td>
<td>155 (29)</td>
</tr>
<tr>
<td>16+</td>
<td>1012 (30)</td>
<td>0.7 (0.6–0.9)</td>
<td>144 (27)</td>
</tr>
<tr>
<td><strong>Gravidity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primigravid†</td>
<td>976 (29)</td>
<td>1.1 (1.0–1.3)</td>
<td>151 (29)</td>
</tr>
<tr>
<td>Multigravid†</td>
<td>2412 (71)</td>
<td>0.9 (0.6–1.5)</td>
<td>377 (71)</td>
</tr>
<tr>
<td><strong>Prepregnancy body mass index (kg/m^2)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Underweight (&lt;18.5)</td>
<td>188 (6)</td>
<td>1.5 (1.1–2.0)</td>
<td>27 (5)</td>
</tr>
<tr>
<td>Normal weight (18.5–24.9)†</td>
<td>1854 (55)</td>
<td>1.5 (1.1–2.0)</td>
<td>282 (53)</td>
</tr>
<tr>
<td>Overweight (25.0–29.9)</td>
<td>707 (21)</td>
<td>0.9 (0.7–1.1)</td>
<td>105 (20)</td>
</tr>
<tr>
<td>Obese (≥30)</td>
<td>484 (14)</td>
<td>1.1 (0.9–1.3)</td>
<td>90 (17)</td>
</tr>
<tr>
<td><strong>Folic acid exposure B1–P1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily use in B1 and P1</td>
<td>778 (23)</td>
<td>0.9 (0.7–1.1)</td>
<td>127 (24)</td>
</tr>
<tr>
<td>Any use in B1 or P1 (not daily)</td>
<td>915 (27)</td>
<td>0.9 (0.7–1.0)</td>
<td>140 (27)</td>
</tr>
<tr>
<td>No use in B1 or P1†</td>
<td>1697 (50)</td>
<td>0.9 (0.7–1.0)</td>
<td>261 (49)</td>
</tr>
</tbody>
</table>

(Continued)
associated with bilateral CLP (OR 4.2). Bilateral clefting is a more severe form, and a 2-fold higher sibling recurrence risk has been reported among infants with bilateral compared with unilateral cleft lip.\textsuperscript{20} In addition, some evidence suggests that genetic risk factors for clefts may differ by laterality.\textsuperscript{21} Therefore, bilateral clefts might be more likely among persons with a certain gene variant that interacts with tobacco exposure. If smoking (and particularly higher doses of smoking) is more strongly associated with bilateral clefts, this might direct research efforts to identify gene variants related to tobacco detoxification among bilateral cases compared with unilateral cases or controls.

The effect estimates for intermediate smoking exposure across the various subtypes of clefts did not follow a trend consistent with a linear dose-response (Table 4). This might indicate a higher degree of exposure misclassification from the lowest to the middle exposure categories, due to under-reporting of any smoking as well as of the amount of smoking. Alternatively, this could suggest a threshold effect, with maternal smoking having an impact only at higher levels of exposure.

Maternal smoking also might impact birth outcomes differentially for male and female infants. Previous studies have suggested that male infants exposed to maternal smoking are more likely to exhibit fetal distress\textsuperscript{22}; deficits in birth weight and length\textsuperscript{23–25}; and birth defects,\textsuperscript{26} including CLP and CPO.\textsuperscript{27} In the current study, the effects of smoking on CLP were similar for males and females. For CPO, for the risk among females was slightly higher than among males. Thus, our data do not support an excess risk among male infants.

Approximately 4000 compounds have been identified in tobacco smoke, including aromatic amines, which can

\begin{table}
\centering
\caption{(Continued)}
\begin{tabular}{llllll}
\hline
Characteristic & \multicolumn{2}{c}{Controls} & \multicolumn{2}{c}{CLP} & \multicolumn{2}{c}{CPO} \\
& (n = 3390) & (n = 933) & & (n = 528) & & \\
\hline
Alcohol use B1–P3 & & & & & & \\
>1.5 drinks per day & 57 (2) & 19 (2) & 1.2 (0.7–2.1) & 11 (2) & 1.3 (0.6–2.5) \\
\leq 1.5 drinks per day & 1248 (37) & 346 (37) & 1.0 (0.9–1.2) & 200 (38) & 1.1 (0.9–1.3) \\
No alcohol use reported\textsuperscript{1} & 2058 (61) & 562 (60) & & 311 (59) & & \\
First-degree family history of orofacial cleft & & & & & & \\
Yes & 14 (0) & 52 (6) & 14.2 (7.6–27) & 28 (5) & 13.5 (6.8–27) & \\
No & 3376 (100) & 881 (94) & & 500 (95) & & \\
Periconceptional maternal smoking\textsuperscript{1} & & & & & & \\
Exposed & 684 (20) & 230 (25) & 1.3 (1.1–1.6) & 122 (23) & 1.2 (1.0–1.5) \\
Unexposed\textsuperscript{1} & 2145 (63) & 552 (59) & 0.6 (0.2–2.0) & 317 (60) & & \\
Unclear (includes ETS exposure) & 561 (17) & 151 (16) & n/a & 89 (17) & n/a & \\
Patterns of maternal smoking & & & & & & \\
Preconception only (B1) & 79 (3) & 28 (4) & 1.4 (0.9–2.1) & 10 (2) & 0.9 (0.4–1.7) \\
Pre- and postconception\textsuperscript{1} & 585 (21) & 199 (25) & 1.3 (1.1–1.6) & 111 (25) & 1.3 (1.0–1.6) \\
Postconception only & 20 (1) & 3 (0) & 0.6 (0.2–2.0) & 1 (0) & 0.3 (0.1–2.5) & \\
Unexposed\textsuperscript{1} & 2145 (76) & 552 (71) & & 317 (72) & & \\
Periconceptional ETS exposure among nonsmokers only & & & & & & \\
ETS at home or work (B1–P3) & 554 (20) & 147 (21) & 1.0 (0.8–1.3) & 88 (22) & 1.1 (0.8–1.4) & \\
Unexposed–ETS (B1–P3)\textsuperscript{1} & 2145 (79) & 552 (79) & 2.0 & 317 (78) & & \\
Patterns of ETS exposure among nonsmokers only & & & & & & \\
ETS preconception only (B1) & 21 (1) & 7 (1) & 1.3 (0.5–3.1) & 7 (2) & 2.3 (1.0–5.3) \\
ETS pre- and postconception\textsuperscript{1} & 494 (18) & 131 (19) & 1.0 (0.8–1.3) & 77 (19) & 1.1 (0.8–1.4) \\
ETS postconception only & 39 (1) & 9 (1) & 0.9 (0.4–1.9) & 4 (1) & 0.7 (0.2–2.0) & \\
Unexposed\textsuperscript{1} & 2145 (79) & 552 (79) & & 317 (78) & & \\
\hline
\end{tabular}
\end{table}

CLP indicates cleft lip with or without cleft palate; CPO, cleft palate alone; B1, month before pregnancy; P1, first month of pregnancy; P2, second month of pregnancy; P3, third month of pregnancy.

\textsuperscript{*} Some totals do not equal 100\% due to missing values and rounding.

\textsuperscript{1} Reference category.

\textsuperscript{2} Mothers who reported smoking in the month before pregnancy or in the first, second, or third month of pregnancy are classified as exposed; mothers with no reported active or passive smoke exposure in the same time period are classified as unexposed; mothers who had passive smoke exposure or who did not answer the questions on smoking in pregnancy are classified as “unclear.”

\textsuperscript{3} Mother reported smoking in B1 and in at least one month of the P1–P3 period.
A recent study found an association of polymorphic variants of NAT1 (a gene that codes for N-acetyl transferase 1, a key enzyme associated with aromatic amine biotransformation in the first trimester) and the risk of orofacial clefts associated with maternal smoking.\textsuperscript{30} Maternal exposure to cadmium, another component of cigarette smoke, has been associated with cleft palate in animal models.\textsuperscript{31,32} Another possible mechanism is that the

### TABLE 3. Reported Maternal Smoking any Time in the Periconceptional Period From 1 Month Before Pregnancy Through the End of the First Trimester

<table>
<thead>
<tr>
<th>Category</th>
<th>Exposed Cases</th>
<th>Exposed Controls</th>
<th>Unexposed Cases</th>
<th>Unexposed Controls</th>
<th>Adjusted OR* (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All orofacial clefts</td>
<td>333</td>
<td>679</td>
<td>819</td>
<td>2136</td>
<td>1.2 (1.0–1.5)</td>
</tr>
<tr>
<td>CLP</td>
<td>218</td>
<td>679</td>
<td>519</td>
<td>2136</td>
<td>1.3 (1.0–1.6)</td>
</tr>
<tr>
<td>Bilateral</td>
<td>53</td>
<td>679</td>
<td>104</td>
<td>2136</td>
<td>1.7 (1.2–2.6)</td>
</tr>
<tr>
<td>Unilateral</td>
<td>136</td>
<td>679</td>
<td>344</td>
<td>2136</td>
<td>1.2 (0.9–1.5)</td>
</tr>
<tr>
<td>Isolated</td>
<td>200</td>
<td>679</td>
<td>447</td>
<td>2136</td>
<td>1.4 (1.1–1.7)</td>
</tr>
<tr>
<td>With multiple defects</td>
<td>18</td>
<td>679</td>
<td>72</td>
<td>2136</td>
<td>0.8 (0.4–1.4)</td>
</tr>
<tr>
<td>Males only</td>
<td>148</td>
<td>345</td>
<td>331</td>
<td>1043</td>
<td>1.3 (1.0–1.7)</td>
</tr>
<tr>
<td>Females only</td>
<td>69</td>
<td>333</td>
<td>187</td>
<td>1090</td>
<td>1.2 (0.8–1.7)</td>
</tr>
<tr>
<td>Isolated, Bilateral</td>
<td>47</td>
<td>679</td>
<td>85</td>
<td>2136</td>
<td>1.8 (1.2–2.8)</td>
</tr>
<tr>
<td>Isolated, Unilateral</td>
<td>131</td>
<td>679</td>
<td>308</td>
<td>2136</td>
<td>1.3 (1.0–1.6)</td>
</tr>
<tr>
<td>CPO</td>
<td>115</td>
<td>679</td>
<td>300</td>
<td>2136</td>
<td>1.2 (0.9–1.5)</td>
</tr>
<tr>
<td>With Pierre Robin sequence</td>
<td>27</td>
<td>679</td>
<td>79</td>
<td>2136</td>
<td>1.1 (0.6–1.8)</td>
</tr>
<tr>
<td>Without Pierre Robin sequence</td>
<td>88</td>
<td>679</td>
<td>221</td>
<td>2136</td>
<td>1.2 (0.9–1.6)</td>
</tr>
<tr>
<td>Isolated</td>
<td>92</td>
<td>679</td>
<td>245</td>
<td>2136</td>
<td>1.2 (0.9–1.6)</td>
</tr>
<tr>
<td>With multiple defects</td>
<td>23</td>
<td>679</td>
<td>55</td>
<td>2136</td>
<td>1.1 (0.6–2.0)</td>
</tr>
<tr>
<td>Males</td>
<td>60</td>
<td>345</td>
<td>137</td>
<td>1043</td>
<td>1.2 (0.8–1.7)</td>
</tr>
<tr>
<td>Females</td>
<td>55</td>
<td>333</td>
<td>162</td>
<td>1090</td>
<td>1.2 (0.8–1.7)</td>
</tr>
</tbody>
</table>

\textsuperscript{CLP} indicates cleft lip with or without cleft palate; CPO, cleft palate alone.

*Adjusted for folic acid use (any use in the month before or first month of pregnancy versus no use), sex, obesity (body mass index \(\geq 30\)), periconceptional alcohol use, maternal education (0–12 yr, >12 yr), and gravidity (primigravid, multigravid). Infants with a first-degree family history of orofacial clefts are excluded.

### TABLE 4. Reported Maternal Smoking by Highest Level of Reported Smoking* in the Periconceptional Period From One Month Before Pregnancy Through the End of the First Trimester

<table>
<thead>
<tr>
<th>Category</th>
<th>Light</th>
<th>Medium</th>
<th>Heavy</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Cases</td>
<td>No. of Cases</td>
<td>No. of Cases</td>
<td>No. of Cases</td>
</tr>
<tr>
<td>Adjusted OR* (95% CI)</td>
<td>Adjusted OR* (95% CI)</td>
<td>Adjusted OR* (95% CI)</td>
<td>Adjusted OR* (95% CI)</td>
</tr>
<tr>
<td>All orofacial clefts</td>
<td>220</td>
<td>81</td>
<td>31</td>
</tr>
<tr>
<td>CLP</td>
<td>147</td>
<td>52</td>
<td>19</td>
</tr>
<tr>
<td>Bilateral</td>
<td>33</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>Unilateral</td>
<td>96</td>
<td>31</td>
<td>9</td>
</tr>
<tr>
<td>Isolated</td>
<td>138</td>
<td>47</td>
<td>15</td>
</tr>
<tr>
<td>With multiple defects</td>
<td>9</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>CPO</td>
<td>73</td>
<td>29</td>
<td>12</td>
</tr>
<tr>
<td>With Pierre Robin sequence</td>
<td>16</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Without Pierre Robin sequence</td>
<td>57</td>
<td>23</td>
<td>7</td>
</tr>
<tr>
<td>Isolated</td>
<td>57</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>With multiple defects</td>
<td>16</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

\textsuperscript{CLP} indicates cleft lip with or without cleft palate; CPO, cleft palate alone.

*Adjusted for folic acid use (any use in the month before or first month of pregnancy versus no use), sex, maternal age (\(\leq 30\) yr, \(>30\) yr), race/ethnicity (non-Hispanic white, non-Hispanic black, Hispanic, other), study site, prepregnancy obesity (body mass index \(\geq 30\), body mass index <30\), periconceptional alcohol use, maternal education (0–12 yr, >12 yr), and gravidity (primigravid, multigravid). Infants with a first-degree family history of orofacial clefts are excluded.
risk could be mediated through folate levels, since some studies have shown that folic acid-containing vitamins decrease the risk for clefts, and one study has shown an interaction between smoking and multivitamin use, with the effect for smoking most evident among mothers who did not take multivitamins. Both active smokers and nonsmokers with high ETS exposure have lower serum folate and red blood cell folate levels than nonsmokers. However, in our data, periconceptional folic acid intake did not markedly change the association between smoking and clefts. Another possible mechanism is related to maternal hypoxia, which has been shown to increase the risk for cleft lip in certain strains of mice. Vasoconstriction of fetal and maternal blood vessels, caused by nicotine and hypoxia, could also result in an increased risk for orofacial clefts, mediated by a decreased supply of nutrients to embryonic tissues.

In this study, ETS exposure was associated with CPO in the presence of multiple defects, with no overall impact of ETS on the occurrence of orofacial clefts. Both associations might simply reflect chance findings. One limitation was our inability to quantify ETS exposure. Studies using cotinine levels to measure dose of ETS exposure have documented adverse reproductive outcomes at the highest levels of ETS exposure.

The validity of the information reported by mothers for both active smoking and ETS exposure is of concern, both because of the time interval between the exposure (early pregnancy) and the exposure assessment (6 weeks to 24 months after the estimated date of delivery), and because of the social undesirability of smoking during pregnancy. Despite this limitation, studies that have compared repeated measures of self-reported smoking during pregnancy have demonstrated specificity of reported nonsmoking at 94% or greater. Comparing reported maternal smoking during pregnancy with serum or urine cotinine levels has demonstrated a high validity for maternal interview smoking reports, with interview reports of smoking confirmed for 85% and nonsmoking confirmed for 95% of pregnant women who enrolled in a study in the United States in 1992. However, some recent studies have suggested that accuracy of reporting is significantly lower among populations of lower socioeconomic status.

Recall bias might have played a role if case-mothers recalled and reported their exposures more completely than control-mothers, and residual confounding could remain from unmeasured confounders. Although the participation rate of the study is relatively high for this type of study (76% for case-mothers and 69% for control-mothers), nonparticipants differed from participants, limiting the generalizability of our findings. Finally, some diagnostic information, such as Pierre Robin sequence, may be inconsistently documented in medical records.

This analysis provides confirmatory evidence of the adverse effects of maternal smoking on infants, which adds to the previously documented effects of increasing infant mortality, stillbirth, and preterm delivery. Although the effect estimates are weak for low levels of maternal smoking, future gene-environment studies will likely identify subgroups of the population at highest risk from smoking during pregnancy. The costs of maternal smoking are substantial and must be measured both in terms of the direct health care costs and by the resulting morbidity and mortality. These results support the importance of smoking prevention and cessation programs among all women of childbearing potential. Additional research is needed to determine how best to communicate to reproductive-aged women the risk of orofacial clefts associated with smoking. However, this effort might actually serve to strengthen existing efforts to reduce smoking during pregnancy because women could be more receptive to messages about the risk of orofacial clefts. The consistency of findings for orofacial clefts and smoking suggest an opportunity for prevention of these serious defects.
REFERENCES


Prepregnancy Body Mass Index and the Occurrence of Severe Hypertensive Disorders of Pregnancy

Lisa M. Bodnar,*†‡ Janet M. Catov,*†‡ Mark A. Klebanoff,§ Roberta B. Ness,*†‡ and James M. Roberts*†‡

Background: Prepregnancy overweight is a risk factor for mild preeclampsia and mild transient hypertension of pregnancy. Its association with severe subtypes of these disorders has received less attention.

Methods: To assess the association of prepregnancy body mass index (BMI) with severe and mild preeclampsia and transient hypertension of pregnancy, we used data from a 1958–1964 prospective cohort study of 38,188 pregnant women receiving care at 12 U.S. hospitals.

Results: There was a monotonic, dose–response relation between prepregnancy BMI and risk of both severe and mild preeclampsia, as well as the risk of severe and mild transient hypertension of pregnancy. Compared with white women with a BMI of 20, the odds ratios for severe preeclampsia at BMI values of 25 and 30 in white women were 1.7 (95% confidence interval = 1.1–2.5) and 3.4 (2.1–5.6), respectively, and 2.1 (1.4–3.2) and 3.2 (2.1–5.0) in black women. The effect of BMI on risk of severe preeclampsia was similar to its effect on mild disease. Compared with the same referent, odds ratios for severe transient hypertension of pregnancy at BMI values of 25 and 30 in white women were 3.6 (2.0–6.5) and 8.8 (4.4–18), respectively, and 3.0 (1.6–5.8) and 4.9 (2.5–9.6) in black women. Overweight was a stronger risk factor for severe than for mild transient hypertension.

Conclusions: Incidence of both mild and severe hypertensive disorders of pregnancy rises with increasing BMI. Escalating obesity rates may increase pregnancy hypertensive disorders and ensuing perinatal morbidity.

(Epidemiology 2007;18: 234–239)
hospitals in the United States were enrolled into this large prospective study. At enrollment, demographic, socioeconomic, and behavioral information and medical history were collected by in-person interview. Detailed data were also collected at each prenatal visit, during labor and delivery and during the postpartum period.

Over 55,000 pregnancies were included in the study. We selected singleton pregnancies from women who identified their race/ethnicity as non-Hispanic white or non-Hispanic black, gave birth from 20 to 45 weeks’ gestation, and did not have chronic hypertension or elevated blood pressure before 24 weeks’ gestation (n = 43,890). From this sample, we excluded women with implausible height or prepregnancy weight values (n = 236) and women who were missing data on prepregnancy weight or height (n = 3866), preeclampsia or transient hypertension of pregnancy (n = 82), or covariates in the final model (n = 1518). The final analytic sample was 38,188. This sample was similar to the entire cohort with respect to the incidence of severe and mild preeclampsia and transient hypertension of pregnancy, prepregnancy BMI, age, race, socioeconomic status, marital status, smoking status, and parity (data not shown).

Blood pressures were measured at entry and each prenatal visit, during labor and delivery, and postpartum. Korotkoff phase 4 (muffling) or phase 5 (disappearance) was used for diastolic blood pressure.25 Urine samples were tested for albumin at each prenatal visit. A validation study in which information on blood pressure and urinary albumin was checked against that in the original medical records showed a high level of accuracy.25

We applied contemporary definitions6 of preeclampsia and transient hypertension of pregnancy to measurements of blood pressure and urinary protein taken at the time of the study. Preeclampsia was defined as gestational hypertension and proteinuria, and return of abnormalities to normal in the postpartum period.6 Gestational hypertension was defined as 2 or more measurements of systolic blood pressure ≥140 mmHg and/or diastolic blood pressure ≥90 mmHg for the first time after 24 weeks of gestation. In the intrapartum period, the first 5 pressures obtained after hospital admission for delivery were averaged. Proteinuria was defined as 2 random urine dipsticks of 1+ protein or one dipstick of 2+ protein. Cases of preeclampsia were considered severe if they had at least one of the following symptoms: systolic blood pressure ≥160 mmHg, diastolic blood pressure ≥110 mmHg, proteinuria of 5 g/24 hours, proteinuria of 3+ or more, oliguria, pulmonary edema, or convulsions/eclampsia. All other cases were considered mild.

Transient hypertension of pregnancy was defined as gestational hypertension in the absence of proteinuria, followed by a return to normal blood pressure postpartum.6 Cases of transient hypertension of pregnancy were considered severe if they had at least one measurement of systolic blood pressure ≥160 mmHg or diastolic blood pressure ≥110 mm Hg. All other cases were considered mild. This definition has been used previously,15,16 though in past papers this entity was termed “severe gestational hypertension.”

Prepregnancy BMI (weight [kg]/height [m]^2) was based on maternal self-report at enrollment of height and prepregnancy weight. BMI was categorized as underweight (BMI <18.5), normal weight (BMI = 18.5–24.9), overweight (BMI = 25.0–29.9), and obese (BMI = ≥30.0). Race/ethnicity was categorized as white or black. Data on marital status (unmarried, married), and smoking status at entry (smoker, nonsmoker) were available. Education, occupation, and family income data were combined into a composite socioeconomic status score (range 1 to 5).27

Analysis

Multinomial logistic regression was used to model the outcome (eg, mild preeclampsia, severe preeclampsia, no preeclampsia) as a function of prepregnancy BMI and the following potential confounders, which were selected a priori from the literature: maternal age (<20, 20–29, ≥30 years), race (black, white), socioeconomic status (range = 1 to 5), marital status (married, unmarried), parity (primiparous, multiparous), height (continuous), and smoking status (smoker, nonsmoker at entry). Separate models were run for preeclampsia and transient hypertension of pregnancy. BMI was curvilinear in the logit of both outcomes. Therefore, it was specified as a series of spline terms in both models. After testing many splines, a restricted quadratic spline with knots at 20 and 30 was determined to have the best fit for both models. Although odds ratios can be calculated for any BMI value along the continuum, we chose 17, 20, 25, 30, and 35 as representative values because they span the BMI distribution and include the knots (the inflection points in the risk curve). Alternative specifications of covariates had no meaningful impact on the results.

Effect modification by race/ethnicity was assessed using a likelihood ratio test (α = 0.10) comparing models with and without BMI-race/ethnicity interaction terms. Because race/ethnicity met our criterion for effect modification in both models, interaction terms were included, with the referent being white women with BMI of 20. The homogeneity test of equivalence was used to determine if the BMI effect estimate differed across the mild and severe outcome subtypes in the multinomial logistic model.28 An alpha of 0.10 was used to reject the null hypothesis that the BMI effect was homogeneous across subtypes.

There were 6991 observations that represented repeated pregnancies to the same woman in the study. We compared results obtained when both models were fitted with and without accounting for clustering on patient identification number. There were no meaningful differences in results for mild preeclampsia or mild transient hypertension. For each of the severe subtypes, the point estimates remained the same and confidence intervals widened when ignoring the clustering. As has been done previously,19,29,30 we chose to ignore the clustering for 2 reasons. First, it had little impact on our conclusions. Second, use of clustering would have precluded us from conducting likelihood ratio tests needed for testing effect modification and the homogeneity test of equivalence.
RESULTS

Preeclampsia occurred in 3.6% of the total population, and was more common in blacks (4.5%) than whites (2.7%) (Table 1). Of all preeclampsia cases, 22% in blacks and 18% in whites were severe. The incidence of transient hypertension of pregnancy was 7.5% in the total population, 7.8% in blacks, and 7.2% in whites. The proportions of transient hypertension cases classified as severe were similar in white (4.5%) and black (4.4%) women. A majority of women in both racial/ethnic groups were normal weight before pregnancy. Black women were more likely than white women to be overweight or obese before pregnancy; they were also more likely to be 21 years of age or younger, unmarried, of low socioeconomic status, and nonsmokers. Black women entered care later in pregnancy and attended fewer prenatal visits than white women.

Among white and black women, there was a striking monotonic, dose-response relation between prepregnancy BMI and the risk of both mild and severe preeclampsia (Fig. 1). The risk of severe preeclampsia in both black and white women was about 2-fold for a BMI of 25 and 3-fold for a BMI of 30 compared with white women who had a BMI of 20 (Table 2). At a BMI of 35, the risk of severe preeclampsia was at least 5-fold greater in both racial/ethnic groups compared with the reference group, although point estimates were imprecise due to a small sample of morbidly obese women. The effect of BMI on the risk of severe preeclampsia was similar to its effect on mild disease (homogeneity test of equivalence, \( P < 0.28 \)). We observed racial/ethnic differences at the extremes of the BMI distribution (Fig. 1; Table 2). A BMI of 17 was protective against both severe and mild preeclampsia in white but not black women. Furthermore the increase in risk of both severe and mild preeclampsia among women with a BMI of 35 was slightly less among black than among white women. The difference between black and white women in the BMI-preeclampsia association met our definition of effect modification (\( P < 0.08 \)).

The risk of the mild and severe subtypes of transient hypertension of pregnancy also increased in a monotonic dose-response manner with prepregnancy BMI in both racial/ethnic groups (Fig. 2, Table 3). However, for both mild and severe transient hypertension, black women had a higher incidence and greater adjusted odds ratios than white women at BMI values less than 25, but beyond 25, their incidence and adjusted odds ratios were lower than whites (Fig. 2; Table 3). The race difference in the BMI-transient hypertension association was statistically significant (\( P < 0.01 \)).

<table>
<thead>
<tr>
<th>TABLE 1. Characteristics of the Study Population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Whites (n = 19,135)</td>
</tr>
<tr>
<td>Blacks (n = 19,053)</td>
</tr>
<tr>
<td>Preeclampsia; %</td>
</tr>
<tr>
<td>Mild</td>
</tr>
<tr>
<td>2.2</td>
</tr>
<tr>
<td>3.5</td>
</tr>
<tr>
<td>Severe</td>
</tr>
<tr>
<td>0.5</td>
</tr>
<tr>
<td>1.0</td>
</tr>
<tr>
<td>Transient hypertension of pregnancy; %</td>
</tr>
<tr>
<td>Mild</td>
</tr>
<tr>
<td>6.8</td>
</tr>
<tr>
<td>7.5</td>
</tr>
<tr>
<td>Severe</td>
</tr>
<tr>
<td>0.3</td>
</tr>
<tr>
<td>0.3</td>
</tr>
<tr>
<td>Prepregnancy BMI (kg/m²); mean ± SD</td>
</tr>
<tr>
<td>22.1 ± 3.7</td>
</tr>
<tr>
<td>22.7 ± 4.1</td>
</tr>
<tr>
<td>Prepregnancy BMI (kg/m²); %</td>
</tr>
<tr>
<td>&lt;18.5</td>
</tr>
<tr>
<td>10.4</td>
</tr>
<tr>
<td>9.6</td>
</tr>
<tr>
<td>18.5–24.9</td>
</tr>
<tr>
<td>73.9</td>
</tr>
<tr>
<td>68.0</td>
</tr>
<tr>
<td>25.0–29.9</td>
</tr>
<tr>
<td>11.8</td>
</tr>
<tr>
<td>16.2</td>
</tr>
<tr>
<td>30.0–34.9</td>
</tr>
<tr>
<td>3.0</td>
</tr>
<tr>
<td>4.6</td>
</tr>
<tr>
<td>≥35.0</td>
</tr>
<tr>
<td>1.0</td>
</tr>
<tr>
<td>1.5</td>
</tr>
<tr>
<td>Maternal age (yrs); %</td>
</tr>
<tr>
<td>&lt;20</td>
</tr>
<tr>
<td>17.3</td>
</tr>
<tr>
<td>34.3</td>
</tr>
<tr>
<td>20–29</td>
</tr>
<tr>
<td>64.8</td>
</tr>
<tr>
<td>52.6</td>
</tr>
<tr>
<td>≥30</td>
</tr>
<tr>
<td>17.9</td>
</tr>
<tr>
<td>13.1</td>
</tr>
<tr>
<td>Marital status; %</td>
</tr>
<tr>
<td>Unmarried</td>
</tr>
<tr>
<td>11.7</td>
</tr>
<tr>
<td>38.4</td>
</tr>
<tr>
<td>Married</td>
</tr>
<tr>
<td>88.3</td>
</tr>
<tr>
<td>64.6</td>
</tr>
<tr>
<td>Socioeconomic status; %</td>
</tr>
<tr>
<td>1 (lowest)</td>
</tr>
<tr>
<td>3.0</td>
</tr>
<tr>
<td>12.8</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>19.6</td>
</tr>
<tr>
<td>41.5</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>29.4</td>
</tr>
<tr>
<td>32.5</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>28.7</td>
</tr>
<tr>
<td>11.4</td>
</tr>
<tr>
<td>5 (highest)</td>
</tr>
<tr>
<td>19.4</td>
</tr>
<tr>
<td>1.8</td>
</tr>
<tr>
<td>Parity; %</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>32.5</td>
</tr>
<tr>
<td>30.1</td>
</tr>
<tr>
<td>1 or more</td>
</tr>
<tr>
<td>67.5</td>
</tr>
<tr>
<td>69.9</td>
</tr>
<tr>
<td>Smoking status at entry; %</td>
</tr>
<tr>
<td>Nonsmoker</td>
</tr>
<tr>
<td>46.5</td>
</tr>
<tr>
<td>57.4</td>
</tr>
<tr>
<td>Smoker</td>
</tr>
<tr>
<td>53.5</td>
</tr>
<tr>
<td>42.6</td>
</tr>
<tr>
<td>Trimester at entry; %</td>
</tr>
<tr>
<td>First</td>
</tr>
<tr>
<td>27.9</td>
</tr>
<tr>
<td>10.6</td>
</tr>
<tr>
<td>Second</td>
</tr>
<tr>
<td>48.6</td>
</tr>
<tr>
<td>61.4</td>
</tr>
<tr>
<td>Third</td>
</tr>
<tr>
<td>23.5</td>
</tr>
<tr>
<td>28.0</td>
</tr>
<tr>
<td>Number of prenatal visits*; mean ± SD</td>
</tr>
<tr>
<td>9.6 ± 4.2</td>
</tr>
<tr>
<td>7.6 ± 3.5</td>
</tr>
</tbody>
</table>

*Total number of visits from booking to delivery.
TABLE 2. Association Between Prepregnancy BMI and the Risk of Mild And Severe Preeclampsia in a Multinomial Logistic Regression Model (n = 35,422)*

<table>
<thead>
<tr>
<th>BMI</th>
<th>Mild Preeclampsia Adjusted† OR (95% CI)</th>
<th>Severe Preeclampsia Adjusted† OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>0.7 (0.6–0.8)</td>
<td>0.8 (0.5–1.1)</td>
</tr>
<tr>
<td>20†</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>25</td>
<td>1.8 (1.5–2.1)</td>
<td>1.7 (1.1–2.5)</td>
</tr>
<tr>
<td>30</td>
<td>3.0 (2.3–3.8)</td>
<td>3.4 (2.1–5.6)</td>
</tr>
<tr>
<td>35</td>
<td>4.9 (3.5–6.9)</td>
<td>7.6 (4.2–13.7)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BMI</th>
<th>Mild Transient Hypertension of Pregnancy Adjusted† OR (95% CI)</th>
<th>Severe Transient Hypertension of Pregnancy Adjusted† OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>20†</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>25</td>
<td>1.8 (1.6–2.0)</td>
<td>3.6 (2.0–6.5)</td>
</tr>
<tr>
<td>30</td>
<td>3.0 (2.6–3.5)</td>
<td>8.8 (4.4–17.6)</td>
</tr>
<tr>
<td>35</td>
<td>4.9 (4.0–6.1)</td>
<td>17.4 (8.5–35.9)</td>
</tr>
</tbody>
</table>

*Women with transient hypertension of pregnancy (n = 2766) were excluded from this model. BMI was specified as a restricted quadratic spline with knots at 20 and 30. Homogeneity test of equivalence to test the null hypothesis that the effect of prepregnancy BMI is the same for severe preeclampsia as for mild preeclampsia in the multinomial logistic regression model, $\chi^2 = 2.5, P = 0.28$. Exclusion of women who entered care in the third trimester had no meaningful impact on the results.

†Adjusted for maternal age, smoking status, marital status, socioeconomic status, parity, and maternal height.

‡Reference category.

FIGURE 2. Association between prepregnancy BMI and the unadjusted prevalence of mild and severe transient hypertension of pregnancy by race/ethnicity. Curves were estimated by calculating predicted probabilities based on an unadjusted multinomial logistic regression model.

effect of BMI was also different for mild and severe transient hypertension of pregnancy (homogeneity test of equivalence $P < 0.01$). Inspection of the adjusted odds ratios suggested that high BMI was a stronger risk factor for severe than for mild transient hypertension.

We found no meaningful differences in our results for preeclampsia or transient hypertension of pregnancy after excluding women who entered care in the third trimester.

TABLE 3. Association Between Prepregnancy BMI and the Risk of Mild and Severe Transient Hypertension of Pregnancy in a Multinomial Logistic Regression Model (n = 36,936)*

<table>
<thead>
<tr>
<th>BMI</th>
<th>Mild Transient Hypertension of Pregnancy Adjusted† OR (95% CI)</th>
<th>Severe Transient Hypertension of Pregnancy Adjusted† OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>20†</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>25</td>
<td>1.8 (1.6–2.0)</td>
<td>3.6 (2.0–6.5)</td>
</tr>
<tr>
<td>30</td>
<td>3.0 (2.6–3.5)</td>
<td>8.8 (4.4–17.6)</td>
</tr>
<tr>
<td>35</td>
<td>4.9 (4.0–6.1)</td>
<td>17.4 (8.5–35.9)</td>
</tr>
</tbody>
</table>

*Women with preeclampsia (n = 1258) were excluded from this model. BMI was specified as a restricted quadratic spline with knots at 20 and 30. Homogeneity test of equivalence to test the null hypothesis that the effect of prepregnancy BMI is the same for mild transient hypertension of pregnancy as for severe transient hypertension of pregnancy in the multinomial logistic regression model, $\chi^2 = 11.4, P < 0.01$. Exclusion of women who entered care in the third trimester had no meaningful impact on the results.

†Adjusted for maternal age, smoking status, marital status, socioeconomic status, parity, and maternal height.

‡Reference category.

DISCUSSION

Our results confirm previous findings of a strong, monotonic dose–response relation between prepregnancy BMI and the risk of mild preeclampsia and mild transient hypertension of pregnancy, and extend previous findings to suggest that a similar dose-dependent association is evident for the severe subtypes of these disorders among both white and black women. Further, our results demonstrate that high BMI is as strong risk factor for severe and mild preeclampsia and severe transient hypertension of pregnancy among both black and white women.

Our findings are consistent with one previous investigation20 that showed a 3.5-fold (95% CI = 1.7–7.5) greater risk of severe preeclampsia among women with a prepregnancy BMI $\geq 32.3$ kg/m$^2$ compared with women who had a BMI $< 32.3$ kg/m$^2$. Other investigators have found that prepregnancy BMI is an important predictor of severe preeclampsia.25 Our results conflict with 2 studies12,21 that reported no association between maternal weight and severe preeclampsia. However, none of these studies used a measure of prepregnancy adiposity. One study12 assessed maternal BMI at 13 to 21 weeks’ gestation, but this measure is confounded by gestational weight gain and by contribution of intravascular and extravascular fluid increases to maternal weight, which are near maximal at this time. The other investigation21 examined maternal weight at an unspecified time before 18 weeks’ gestation, and did not adjust for height. We are aware of only one study22 that assessed prepregnancy...
BMI in relation to severe gestational hypertension. Investigators reported that BMI was positively associated with “severe hypertension,” but they did not eliminate women with pre-existing hypertension from the case group, nor did they report any effect estimates to allow one to evaluate the strength or precision of the association.22

The mechanisms explaining why maternal overweight predisposes women to mild and severe hypertensive disorders of pregnancy have not been well studied. In a cohort of predominantly mild preeclampsia cases, we previously showed that the strong linear relation between prepregnancy BMI and preeclampsia risk was partially mediated by C-reactive protein, a marker of systemic inflammation, and triglycerides.10 Other important mediators may be oxidative stress, insulin resistance, endothelial dysfunction, reduced immune function, other markers of dyslipidemia, or lifestyle factors such as poor prenatal diet and prenatal physical inactivity. Research is needed to delineate the relevance of these or other pathways in the development of severe forms of transient hypertension of pregnancy and preeclampsia in overweight women, and whether the pathways vary by race/ethnicity.

One of our most intriguing findings was the interaction between BMI and race/ethnicity. Lean black women had a higher risk of preeclampsia and transient hypertension of pregnancy than lean white women, but at high BMI values, the trend was reversed: overweight black women had a lower risk of these disorders than overweight white women. While the effect modification by race met our a priori definition for both outcomes, the ethnic differences were more striking for transient hypertension of pregnancy. For preeclampsia, the differential effect of ethnicity was restricted to the tails of the BMI distribution.

To our knowledge, the interaction between overweight and race/ethnicity on the risk of transient hypertension of pregnancy has not been investigated previously, and only one study assessed this interaction in relation to preeclampsia. In a prospective cohort study of 22,658 women, investigators reported that the effect of obesity (defined as a BMI >29 kg/m²) on preeclampsia risk was slightly lower among black women (adjusted OR = 1.6; 95% CI = 1.1–2.4) than white women (2.5; 1.9–3.5).31 However, the results are difficult to interpret because BMI was based on weight at the initial visit, which may have been in mid-to-late gestation in some women (no data on gestational age at entry are provided), and therefore is likely confounded by gestational weight gain. Other investigators have suggested that adiposity is a less important predictor of certain nonperinatal health outcomes in black than white women. For instance, a high BMI among black women has a substantially weaker association with all-cause mortality,32 coronary heart disease mortality,33,34 and incident cardiovascular disease35 than among white women.

The BMI–race/ethnicity interaction may be explained by the tendency for black women to have central and abdominal body fat distribution compared with white women.35,36 It is thought that atherogenic risk factors such as dyslipidemia and degree of peripheral insulin resistance are not as greatly influenced by central and abdominal obesity.36 Because hypertensive disorders of pregnancy share many similarities with cardiovascular disease,37 these mechanisms may likely be relevant to our findings. Nevertheless, little is currently known about how racial/ethnic differences in fat patterning relate to preeclampsia and transient hypertension of pregnancy. Given the racial/ethnic disparity in the incidence of these disorders,38,39 studies should assess racial/ethnic interactions with pregravidity adiposity and strive to understand mechanisms underlying these differences.

As with other studies of maternal overweight and hypertensive disorders, our study was limited by the potential for right-censoring. Underweight women were more likely to deliver preterm than heavier women (data not shown), so they had less time to be at risk for severe preeclampsia. This may have biased our results upwards and away from the null. Nevertheless, we found a similar monotonic dose–response association between BMI and risk of mild and severe forms of both hypertensive disorders even after restricting the analysis to women who delivered at term (≥37 weeks). Because the Collaborative Perinatal Project was conducted in the 1960s, blood pressure cuffs may not have been available in extra-large sizes to fit obese women. Cuffs that were too small would cause falsely elevated blood pressure in obese subjects, leading to differential misclassification and bias away from the null. Nevertheless, this may not be a major problem since we observed similar strong, dose-dependent relations when we eliminated obese women from our analyses (data not shown). We were limited by a relatively small proportion of obese women in this dataset, which caused unstable point estimates at high BMI values. We also lacked data on other surrogate markers of prepregnancy adiposity, such as prepregnancy waist circumference or skinfold measurements, which may be more potent measures of the adiposity phenotype associated with pregnancy hypertensive disorders.

This study had many notable strengths. The Collaborative Perinatal Project afforded us the unique opportunity to prospectively study the rare outcomes of severe preeclampsia and severe transient hypertension of pregnancy separately by race/ethnicity because of the large sample size and the proven validity of the hypertension and proteinuria data.25 While the standard of care for hypertensive disorders of pregnancy in the 1960s, when the study was conducted, differs from present day, these data offered the chance to study the natural progression of the disease. Our use of spline regression makes no unreasonably strong assumptions about the shape of the BMI dose-response curve, and allows for greater flexibility and more biologically plausible and relevant results than linear or categorical models.40,41 Maternal self-report of prepregnancy weight was unlikely to be an important source of bias in our analysis because self-reported weight has been influenced by central and abdominal obesity.36 Because hypertensive disorders of pregnancy share many similarities with cardiovascular disease,37 these mechanisms may likely be relevant to our findings. Nevertheless, little is currently known about how racial/ethnic differences in fat patterning relate to preeclampsia and transient hypertension of pregnancy. Given the racial/ethnic disparity in the incidence of these disorders,38,39 studies should assess racial/ethnic interactions with pregravidity adiposity and strive to understand mechanisms underlying these differences.

Our results suggest that the escalating rates of obesity in U.S. childbearing-aged women may predict an increase in perinatal morbidity and mortality associated with severe hypertensive disorders of pregnancy. A preconceptional weight loss or lifestyle intervention should be explored to study
simultaneously the prevention of hypertensive disorders of pregnancy and other adverse perinatal outcomes related to obesity.\(^5\) Moreover, future studies should explore the mechanisms underlying the effect of overweight on hypertensive disorders of pregnancy so as to aid in the design of appropriate prenatal interventions for women who enter pregnancy with a high BMI.

**REFERENCES**

34. Stevens J, Juhuari, Cai J. Changes in body mass index prior to baseline among participants who are ill or who die during the early years of follow-up. Am J Epidemiol. 2001;153:946–953.
Variation in Incidence of Neurodevelopmental Disorders With Season of Birth

Hjördis Ösk Atladóttir,*† Erik T. Parner,*† Diana Schendel,‡ Søren Dalsgaard,§ Per Hove Thomsen,§ and Poul Thorsen*

Background: The etiologies of autism spectrum disorder and many neurodevelopmental disorders are largely unknown. The detection of a seasonal variation of birth of children diagnosed with a certain disorder could suggest etiological factors that follow a seasonal pattern. We examined the seasonal variation of births of children diagnosed with any of 4 common childhood neuropsychiatric disorders: autism spectrum disorder, hyperkinetic disorder, Tourette syndrome, and obsessive-compulsive disorder.

Methods: The study cohort consisted of all children born in Denmark from 1990 through 1999 identified in the Danish Medical Birth Register (n = 669,995). Outcome data consisted of both inpatient and outpatient diagnoses reported to the Danish National Psychiatric Registry from 1995 through 2004 using the International Classification of Diseases, 10th edition, diagnostic coding system. Logistic regression combined with spline (a smoothing method) was used to estimate the variation with season of birth for each disorder. Estimates of risk of each disorder with season of birth were adjusted for differences in follow-up time and change in incidence over time.

Results: No convincing variations in season of birth were observed for any of the 4 disorders, or for the autism-spectrum-disorder subtypes.

Conclusion: Although we cannot rule out the possibility of seasonal variation of birth for a range of childhood neurodevelopmental disorders, we find little evidence that seasonal environmental factors are related to these disorders.

(Epidemiology 2007;18: 240–245)

The detection of seasonal variation in the birth of children with neuropsychiatric disorders could provide clues to etiological factors that follow a similar seasonal pattern. An excess of winter–spring births has consistently been reported in schizophrenia, leading to speculations of an association between schizophrenia and factors following a seasonal pattern. Such factors could include the mother’s nutritional condition during pregnancy, her exposure to influenza or other infections, maternal hormones, or meteorological factors, such as humidity or temperature. These risk factors could possibly affect the development of other neuropsychiatric disorders.

In this study, we examined the seasonal variation of birth of children diagnosed with any of 4 common neuropsychiatric developmental disorders that emerge in childhood: hyperkinetic disorder, obsessive-compulsive disorder, Tourette syndrome, or autism spectrum disorder. All 4 disorders have specific diagnostic criteria and high comorbidity.

In ICD-10 (International Classification of Diseases, 10th Edition*), autism spectrum disorder includes a group of disorders characterized by qualitative abnormalities in reciprocal social interactions; patterns of communication; or a restricted, stereotyped, repetitive repertoire of interests and activities; or a combination thereof. Hyperkinetic disorder is characterized by lack of persistence in activities that require cognitive involvement and a tendency to move from one activity to another without completing any one, together with disorganized, ill-regulated, and excessive activity. The essential feature of obsessive-compulsive disorder is recurrent obsessive thoughts or compulsive acts. Tourette syndrome is a form of tic disorder in which there are multiple motor tics and one or more vocal tics. We studied the seasonal variation of birth of children diagnosed with one or more of these psychiatric disorders.

METHODS

Study Design

This was a population-based cohort study that included all children born in Denmark from 1 January 1990 through 31 December 1999. The cohort was identified using the Danish Medical Birth Registry. The Registry was established in 1968 and comprises data on all live births and stillbirths by women with permanent residence in Denmark. All live-born children in Denmark are assigned a Central Population Registry number, which is a unique 10-digit number used for all official personal registrations in Denmark since 1968.

Data on outcome were from the Danish National Psychiatric Registry. All inpatient admissions to psychiatric hospitals and psychiatric wards in general hospitals in Den-
mark have been registered since 1969. Psychiatric outpatient contact has been reported since 1 January 1995. The Psychiatric Registry includes data on clinical diagnoses, dates of admission and discharge, and terms of admission. All diagnoses are made by psychiatrists, and the ICD-10 diagnostic coding criteria have been used since 1994. The specific diagnostic codes used for this study are provided in Table 1. For autism spectrum disorder, analyses were performed for the disorder as a whole, as well as for each subdiagnosis.

Data from the Birth Registry were linked to the Psychiatric Registry using the Central Population Registry number. For this study, follow-up time for a reported diagnosis was restricted to the time period after inclusion of both inpatient and outpatient registration and adoption of ICD-10 in the Psychiatric Registry (1 January 1995 through 31 December 2004). The youngest age of reported diagnosis was 3 years for hyperkinetic disorder, 1 year for autism spectrum disorder and subgroups, and 5 years for obsessive compulsive disorder or Tourette syndrome. To insure that no children included in the analyses for a particular disorder received a diagnosis before 1 January 1995, different birth cohorts within the larger study cohort were used for the analysis of each disorder. Table 1 displays the different analytic birth cohorts for each disorder.

This study was approved by the National Board of Health and the Danish Data Protection Agency.

### Analytic Approach

The effect of any changes in incidence of reported diagnoses over time needed to be removed to estimate the actual seasonal variation. For example, increased rates of autism and attention deficit hyperactive disorder have been reported in recent years. Such increases can impose a crude seasonal variation when data are pooled across years. This effect of the time trend on seasonal variation is shown in Figure 1, in which the risk of being diagnosed with hyperkinetic disorder for children born in December is increased compared with children who are born in January. Consequently, we adjusted for time trends in the analysis as described below.

Each disorder was analyzed separately. Children with more than one autism spectrum disorder subdiagnosis were included in the analysis for the most severe autism spectrum disorder subdiagnosis (defined as the diagnosis with the lowest extension number on the F84.x code), as well as in the analysis for the combined group of autism spectrum disorder. Some children had 2 or more different diagnoses, such as Tourette syndrome and hyperkinetic disorder; in those cases, the child was included as an outcome in the analysis for each separate condition. With this approach, the analysis of each condition could include children with only one diagnosis and children with 2 or more different diagnoses. The association with seasonal variation could differ for children with single and multiple conditions. However we could not distinguish these because of the small numbers of eligible cases. However, we did perform a sensitivity analysis by excluding children with multiple diagnoses and reestimating the seasonal variation effect for each condition based only on children with a single diagnosis.

The follow-up time began for all children at birth and continued until the date of a relevant diagnosis or death or until the end of follow-up time on 31 December 2004. No

### TABLE 1. A Description of the Birth Cohort Used for Each Disorder

<table>
<thead>
<tr>
<th>ICD-10 Code</th>
<th>Birth Cohort</th>
<th>No. Cases</th>
<th>Children Born at ≥37 Weeks</th>
<th>No. (%)</th>
<th>No. Births</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyperkinetic disorder</td>
<td>F90</td>
<td>1992–1999</td>
<td>2033</td>
<td>1722 (85)</td>
<td>542,213</td>
</tr>
<tr>
<td>Obsessive-compulsive disorder</td>
<td>F42</td>
<td>1990–1995</td>
<td>485</td>
<td>431 (89)</td>
<td>402,315</td>
</tr>
<tr>
<td>Tourette syndrome</td>
<td>F95.2</td>
<td>1990–1995</td>
<td>259</td>
<td>235 (91)</td>
<td>402,315</td>
</tr>
<tr>
<td>Autism spectrum disorder</td>
<td>F84.0, F84.1, F84.5, F84.8, F84.9</td>
<td>1994–1999</td>
<td>1860</td>
<td>1669 (90)</td>
<td>407,118</td>
</tr>
<tr>
<td>Childhood autism</td>
<td>F84.0</td>
<td>1994–1999</td>
<td>714</td>
<td>630 (88)</td>
<td>407,118</td>
</tr>
<tr>
<td>Atypical autism</td>
<td>F84.1</td>
<td>1994–1999</td>
<td>145</td>
<td>130 (90)</td>
<td>407,118</td>
</tr>
<tr>
<td>Asperger syndrome</td>
<td>F84.5</td>
<td>1994–1999</td>
<td>350</td>
<td>325 (93)</td>
<td>407,118</td>
</tr>
<tr>
<td>Pervasive developmental disorder–not otherwise specified</td>
<td>F84.8, F84.9</td>
<td>1994–1999</td>
<td>651</td>
<td>582 (90)</td>
<td>407,118</td>
</tr>
</tbody>
</table>

FIGURE 1. The seasonal variation of birth for Danish children born 1992–1999 and diagnosed with hyperkinetic disorder, with no adjustments made for the general trend for increased incidence of reported diagnoses over time.
The present analysis included 2 spline functions. The first spline function described the seasonal variation using 6 spline variables, corresponding to 3 knots, equally spaced. More than 4 variables are seldom required in a restricted cubic spline analysis, but because the present study accounts for incomplete follow-up time. However, Cox regression assumes proportionality of the hazards, and our data did not meet this assumption. When we compared our cruder adjustment for right-censored data with Cox regression analyses, we found there was no major difference in the results.

Seasonal variation in the birth of diagnosed children was estimated by logistic regression, using STATA version 8.0 (StataCorp, College Station, TX). The risk of being diagnosed depended on the length of the follow-up time, possible increase in risk over time (time trends), and possible seasonal variation of births. These 3 factors were described in a model using the cubic spline method. Splines are a smoothing technique, and the cubic spline is a variant that consists of piecewise cubic polynomials. The spline is modeled as a set of explanatory variables (the spline variables) in a regression analysis. The spline function allowed the assumption of a periodic season. A periodic season involved 2 restrictions: (1) the risk estimates at the end of December were restricted to be equal to the risk estimates at the beginning of January and (2) the slope of the curve at the end of the year was restricted to be equal to the slope of the curve at the start of the year. Because of these 2 restrictions, the number of parameters ( spline variables) included in the seasonal spline function was reduced by two. The number of knots was, however, unchanged. The validity of the assumption of a periodic season was confirmed using a likelihood ratio test.

The present analysis included 2 spline functions. The first spline function described the seasonal variation using 6 spline variables, corresponding to 3 knots, equally spaced. More than 4 variables are seldom required in a restricted cubic spline analysis, but because the present study included a large sample size the seasonal variation was fitted using 6 variables. The assumption of a periodic season reduced the number of spline variables to four. Thus only 4 degrees of freedom were used in estimating the seasonal variation and in testing the statistical significance of the seasonal variation; hence, increased statistical power was obtained. The second spline adjusted for the time trend and difference in follow-up time, ie, the second spline described calendar time. This second spline function was fitted using 3 spline variables. The reason for using a relatively low number of spline variables in the second spline was to obtain a rather smooth spline function and thus to avoid coincidently adjusting for a part of the seasonal variation. The final analysis was performed using logistic regression on 7 spline variables. The test for a seasonal variation of birth was performed by testing the statistical significance of the seasonal spline function (the 4 seasonal spline variables), ie, hypothesizing no seasonal variation of birth.

The use of the spline variables in the analysis allowed a day-by-day analysis of the seasonal variation of birth and thereby avoided grouping the year into months, quarters, or other arbitrary time intervals within the year, as has been done in previous studies. The relative risk (RR) curves compared the risk of a particular day to the average risk over the whole year. The 95% confidence interval (CI) was calculated according to the fit of the spline model, ie, the confidence interval corresponded to the fit evaluated for each individual day. To demonstrate the adequacy of the fit of the smoothed seasonal-spline function, we display in Figure 2 the point estimates obtained by grouping the season into months, while still adjusting for the time trend and differences in follow-up time using the second spline.

Seasonal variation could be associated with risk factors arising at different times in pregnancy, such as during the first or second trimester, and not just around the date of delivery. If variation in length of pregnancy was not considered, then seasonal variation in pregnancy exposures arising at times other than delivery might be missed. Alternatively, variation in gestational length might be on the causal pathway in the association between season of birth and outcome, such as if seasonal variation in infectious exposures leads to preterm births, which in turn could lead to one of the studied neurodevelopmental outcomes. Although it was desirable to explore the relation between season of birth and outcome by gestational age, the numbers of preterm (<37 weeks’ gestation) children for obsessive-compulsive disorder, Tourette syndrome, or the autism spectrum disorder subdiagnoses were generally too small for reliable estimates; subanalyses of preterm children were performed only for children with hyperkinetic disorder or autism spectrum disorder as a whole. For all diagnoses, to control for differences in length of gestation, we performed subanalyses of the relation between season of birth and outcome that included only term (>37 weeks’ gestation). For each disorder, the number of children born at term is displayed in Table 1.

Risk factors arising at the exact time of conception could be of great importance. Data included information on gestational age, and so we were able to estimate date of conception. Therefore, we additionally performed the analyses based on estimates of the season of conception, rather than season of birth.

RESULTS

A total of 4376 children were diagnosed with a total of 4637 disorders. The numbers of children diagnosed with each disorder are shown in Table 1.

Figure 2 shows the adjusted seasonal variation in birth for the various subdiagnoses. The figure displays the seasonal variation from the smoothing function, with 95% CI and the adequacy of the fit of the smoothing function, by demonstrating risk estimates after grouping the birth year into months. For each of the disorders, the relative risk estimates for birth at different dates during the year ranged from slightly under 1.0 to slightly over 1.0. The CIs were calculated separately for each day; CIs can be used only to estimate the uncertainty of the measurement of each day, not to compare 2 separate days. Atypical autism had the greatest variation in relative risk estimates, ranging from the lowest risk estimate of 0.71 (95% CI = 0.49–1.02) for children born in September, to the highest risk estimate of 1.29 (0.90–1.85) for children born in
December. Autism spectrum disorder overall had the least variation in risk estimates, ranging from 0.96 (0.89–1.04) for children born in May to 1.06 (0.95–1.18) for children born in February. Disorders with a relatively low case count, such as atypical autism and Tourette syndrome, displayed a great variation in the range of risk estimates within a year, while disorders with a high case count, such as hyperkinetic disorder or the autism spectrum disorder overall, displayed only minor variation in risk estimates within the year.

No convincing evidence for variation in season of birth was observed for any of the 4 disorders, or the autism spectrum disorder subdiagnoses (hyperkinetic disorder: $P = 0.06$; Tourette syndrome: $P = 0.14$; obsessive compulsive disorder: $P = 0.71$; autism spectrum disorder: $P = 0.83$; childhood autism: $P = 0.52$; atypical autism: $P = 0.27$; Asperger syndrome: $P = 0.55$; and pervasive developmental disorder—not otherwise specified: $P = 0.83$). There was no change in these results based on the subanalyses of term children alone, the subanalyses of children born preterm with hyperkinetic disorder or autism spectrum disorder, or the sensitivity analyses of children with a single diagnosis. When analyzing data by season of conception, the association of season with hyperkinetic disorder became slightly stronger ($P = 0.04$).

**DISCUSSION**

This study found no substantial seasonal variation in birth for hyperkinetic disorder, obsessive-compulsive disorder, Tourette syndrome, autism spectrum disorder, or for any one of the autism spectrum disorder subdiagnoses. There was some evidence of a seasonal effect for hyperkinetic disorder, with higher rates in autumn and lower in spring.

Many previous studies have looked at seasonal variation in the birth of children with autism, with inconsistent results. Earlier studies, many with small sample sizes, found an increased risk of developing autism among births in certain months, while more recent and often larger studies have reported no consistent seasonal variation. No studies have explored seasonal variations in birth for children diagnosed with hyperkinetic disorder, although a report on attention-deficit/hyperactivity disorder (ADHD) found that certain small subgroups of children with ADHD had an increased risk of being diagnosed with ADHD if they were born in September. Diagnostically, ADHD (based on the DSM-IV (Diagnostic and Statistical Manual of Mental Disorders – Fourth Edition) overlaps hyperkinetic disorder (based on the ICD-10) regarding symptoms of hyperactivity,
but, unlike hyperkinetic disorder, ADHD also permits symptoms of inattention in the absence of hyperactivity. Greenberg found no seasonal variation of birth for obsessive compulsive disorder in 1981. No studies have been published on seasonal variation of birth of children with Tourette syndrome.

Previous studies have had a number of limitations which are addressed in the present study. Unlike prior studies, we applied a day-by-day analysis. This approach is preferred, compared with arbitrarily dividing the year into months or quarters. Dividing the year into months causes the loss of statistical power, either by multiple testing in which 12 separate significance tests are performed or by the use of a single test with 12 categories resulting in a test with 11 degrees of freedom. It is not known if the time of birth or the time of different prenatal periods is of importance when studying seasonal variation. Thus, the crude and arbitrary classification of season of birth into months or yearly quarters gives limited information on both the exact time of birth and the time of different periods within the pregnancy. This study tested for seasonal variation of birth using 4 degrees of freedom and no categorization of the season of birth.

Bolton et al emphasized the necessity of adjusting for the general time trend when studying the seasonal variation of birth of children with disorders that have a change in incidence by year of birth. In the present study, the increase in incidence of reported diagnoses by year of birth affected the seasonal variation. The chance of being diagnosed increased continuously from January through December each year. This created an artifact when many continuous years were combined, causing a dramatic decrease in risk for select disorders between births at the end of December (highest risk) and the beginning of January (lowest risk). This risk difference was unlikely to have been associated with etiologic mechanisms linked to seasonal variation because living conditions (e.g., weather, diet, or infection rates) among these 2 consecutive months were similar. Adjusting for time trends in incidence of reported diagnosis removed this artifact, and the true seasonal variation could then be estimated.

This study involved large cohorts of children based on data from a nationwide psychiatric register in which diagnostic information for all conditions were uniformly reported using the same diagnostic coding system (ICD-10). Nevertheless, a limitation of the present study was that, while all diagnoses in the Psychiatric Registry are reported by psychiatrists, generally the diagnoses have not been validated by external review. A small pilot validation of the childhood autism diagnosis verified the validity of 40 infantile autism diagnoses recorded in the Psychiatric Registry; 92% met the criteria for autistic disorder, based on a coding scheme developed by the Centers for Disease Control and Prevention. An unofficial validation of 171 hyperkinetic disorder diagnoses was performed by Linnet in 2004; the agreement percentage on a full diagnosis of ADHD was 89%, while the remaining 11% lacked only one symptom to fulfill this diagnosis. While not comprehensive or definitive, these studies suggest that, in general, the validity of the reported diagnoses in the Psychiatric Registry is acceptable.

Although our putative marker for exposure was date of birth, there could have been other time points in pregnancy at which the true seasonal exposure of risk occurred. We attempted to control for this possibility by limiting our subanalysis to full-term children, so that season of birth also served as a proxy measure for season of conception, first trimester, second trimester, etc. We also attempted to perform the analysis for preterm children because of the possible link between prematurity birth and infection during pregnancy. Due to the limited number of preterm births, this analysis was possible only for hyperkinetic disorder and autism spectrum disorder overall (Table 1). For all disorders, no significant seasonal variation was observed for the subgroups of children born at term, nor was there seasonal variation observed for children diagnosed with autism spectrum disorder or hyperkinetic disorder and born preterm. We combined all preterm births into a single category, although the association between season of birth and outcome could have varied by gestational age within this heterogeneous group. The inability to perform analyses with more narrowly defined ranges of gestational age might have reduced our ability to detect a seasonal variation effect on autism spectrum disorder overall or hyperkinetic disorder among the preterm children. When analyzing data by season of conception, only one disorder (hyperkinetic disorder) reached statistical significance ($P = 0.04$). We believe that caution is required when interpreting this finding. Although this result might represent a true variation in season of conception, it may also represent the result of multiple testing—a mere chance finding. Replication in other data sets would be needed before this single finding could be given more weight.

Seasonal variation of birth is a fairly crude proxy measure for more specific exposures that vary seasonally and may be linked etiologically with one of these disorders. In fact, seasonal exposures that vary in the time of year might counterbalance each other and be undetectable in a general analysis of season of birth. Thus, the lack of a seasonal variation of birth does not exclude the possibility that one or more environmental factors, such as infection, could be related to these disorders.

REFERENCES


Guided Multiple Imputation of Missing Data

Using a Subsample to Strengthen the Missing-at-Random Assumption

Gary Fraser and Ru Yan

Abstract: Multiple imputation can be a good solution to handling missing data if data are missing at random. However, this assumption is often difficult to verify. We describe an application of multiple imputation that makes this assumption plausible. This method requires contacting a random sample of subjects with incomplete data to fill in the missing information, and then adjusting the imputation model to incorporate the new data. Simulations with complete data that were decided not missing at random showed, as expected, that the method restored the original beta coefficients, whereas other methods of dealing with missing data failed. Using a dataset with real missing data, we found that different approaches to imputation produced moderately different results. Simulations suggest that filling in 10% of data that was initially missing is sufficient for imputation in many epidemiologic applications, and should produce approximately unbiased results, provided there is a high response on follow-up from the subsample of those with some originally missing data. This response can probably be achieved if the data collection is planned as an initial approach to dealing with missing data if data are missing at random. However, this assumption may not be warranted.

Multiple imputation is a strategy that uses observed data to impute missing data, ideally when data are “missing at random.” This term designates a missingness pattern such that the probability of a data point being missing depends only on the data that are observed. The target analysis can then proceed incorporating both original observed and imputed data to produce unbiased results. The standard error of such an estimate is underestimated, however, unless further action is taken. Combining results of target analyses from several such partially imputed data sets will, when data are missing at random, produce unbiased coefficient estimates with correct standard errors. The standard errors are of course greater than those that would result from an analysis of an initially complete data set.

A central problem is whether the missing data are in fact missing at random. For instance, if the means of nonmissing values (conditional on covariates) differ from those of the true values of the data of the same variables, these missing data are not missing at random. Yet, it is often difficult to be certain just what has caused the missing data, and on which variables the probability of “missing” may depend.

In this paper, we slightly extend a survey strategy of obtaining more information about missing survey subjects and note the natural application to epidemiologic studies in which some study members have provided incomplete data. This approach strengthens the claim that the missing data are missing at random. The goal is to fill in a random subsample of the initially missing data and adjust the imputation model using this extra information as a guide. Such a subsample survey can usually be completed more successfully among epidemiologic study participants than among nonresponders to a survey.

METHODS

Consider a data set \( Y \) of \( J \) variables on each of \( N \) subjects \( i = 1, \ldots, N \). \( Y \) includes both outcome and predictor variables. For each of the \( J \) variables, \( V(j) \), there is a number \( d(j) \), \( j = 1, \ldots, J \), that are missing. Divide the subjects into those who have missing data on at least one variable, set \( S(1) \), and those with no missing data, set \( S(2) \). Take a random sample \( R \) of subjects in set \( S(1) \) and expend the necessary resources to contact a high proportion of biased results. Others have assumed that subjects who, for instance, skip a dietary question do so because they do not eat that food; a zero is then imputed. However, such an approach may not be warranted.

Multiple imputation is a strategy that uses observed data to impute missing data, ideally when data are “missing at random.” This term designates a missingness pattern such that the probability of a data point being missing depends only on the data that are observed. The target analysis can then proceed incorporating both originally observed and imputed data to produce unbiased results. The standard error of such an estimate is underestimated, however, unless further action is taken. Combining results of target analyses from several such partially imputed data sets will, when data are missing at random, produce unbiased coefficient estimates with correct standard errors. The standard errors are of course greater than those that would result from an analysis of an initially complete data set.

A central problem is whether the missing data are in fact missing at random. For instance, if the means of nonmissing values (conditional on covariates) differ from those of the true values of the data of the same variables, these missing data are not missing at random. Yet, it is often difficult to be certain just what has caused the missing data, and on which variables the probability of “missing” may depend.

In this paper, we slightly extend a survey strategy of obtaining more information about missing survey subjects and note the natural application to epidemiologic studies in which some study members have provided incomplete data. This approach strengthens the claim that the missing data are missing at random. The goal is to fill in a random subsample of the initially missing data and adjust the imputation model using this extra information as a guide. Such a subsample survey can usually be completed more successfully among epidemiologic study participants than among nonresponders to a survey.

METHODS

Consider a data set \( Y \) of \( J \) variables on each of \( N \) subjects \( i = 1, \ldots, N \). \( Y \) includes both outcome and predictor variables. For each of the \( J \) variables, \( V(j) \), there is a number \( d(j) \), \( j = 1, \ldots, J \), that are missing. Divide the subjects into those who have missing data on at least one variable, set \( S(1) \), and those with no missing data, set \( S(2) \). Take a random sample \( R \) of subjects in set \( S(1) \) and expend the necessary resources to contact a high proportion of biased results. Others have assumed that subjects who, for instance, skip a dietary question do so because they do not eat that food; a zero is then imputed. However, such an approach may not be warranted.
subjects in this sample to fill in their missing data. All subjects in R with missing data for any combination of variables that includes missingness for a specific variable, say \( V(j) \), will also be a random sample of those in Y with missing data for that variable.

For each variable \( V(j) \) that has a nonzero \( d(j) \), create dummy variables \( M(i, j) \), which takes values 1 to represent original missingness, otherwise zero. Thus, subjects in set R with originally missing data for \( V(j) \) will have a value of 1 for \( M(i, j) \), but will also have a filled-in value for \( V(j) \). Other subjects in Y but not R with missing data for \( V(j) \) will also have a value of one for \( M(i, j) \) but their \( V(j) \) remains missing.

Then \( pr( M, Y, \theta) = pr( M|Y(\text{obs}), Y(\text{miss}), \theta) = pr( M|Y(\text{obs}), Y(R), \psi, \theta) \) where \( M \) is a missingness indicator set to 1 to indicate initial missing status, otherwise 0; \( Y(\text{obs}) \) is that part of \( Y \) that is observed; \( Y(\text{miss}) \) is that part of \( Y \) that was originally missing; \( Y(R) \) is the part of \( Y(\text{miss}) \) that, although originally missing, is now filled in; M is the matrix of \( M(i, j) \); \( \theta \) is the vector of parameters; and \( \psi \) is the sampling proportion, \( R/S(1) \), for the subsample. This probability no longer depends systematically on the part of \( Y(\text{miss}) \) that continues to be missing, as it is properly represented by \( Y(R) \), a random sample of \( Y(\text{miss}) \). By definition, the remaining missing data are missing at random, assuming correct specification of the likelihood.

After assuming that parameters of the distributions of \( M \) and \( Y \) have distinct spaces, the imputation can proceed considering only the likelihood conditioned on observed data.

In the applications that follow, covariates used to predict missing data are modeled as a linear sum, \( h \), that includes the categorical \( M \) variable. Missing data for continuous variable \( V(j) \), for instance, is imputed by a draw from the probability or density function \( f(h) \), where \( h = \alpha + \sum_{k} \beta_k V_{ik} + \gamma_i M_{ij} + \epsilon_i \) and \( M_{ij} \) is set to 1; and \( j \) depends on the form of the likelihood function for the data. The \( V_{ij} \) may be continuous or categorical in form, and should include the dependent variable of the final regression of interest.

Where the missing \( V_j \) is a categorical variable, the imputation is by a draw from the probability distribution that includes a log-linear component, and a component incorporating a design matrix that models a link between the categorical and continuous variables. We chose to include product terms between all categorical variables (including the \( M \) variables) in the log-linear component. Thus the imputation is guided both by data initially present and data filled in later in the random subsample.

Further details are found in chapter 9 in the book by Schafer. We implemented his mixed model with code available free on the world wide web. Continuous variables are assumed to be normally distributed.

The linear sum, \( h \), implies that imputed values of \( V_j \), conditional on covariates, differ systematically from those observed by only a different constant, \( \gamma_i \), for each variable. More complicated scenarios, such as different scaling coefficients among nonresponders for some variables, could be modeled. However, in an epidemiologic study, often it will be reasonable to assume that those who are missing some data (usually a high percentage of subjects) do not differ systematically from others in the relationships between variables.

The imputation is performed \( Q \) times to provide \( Q \) plausible complete data sets, and the desired analysis is run on each. Then the \( Q \) sets of coefficients and variance estimators are combined as described by Little and Rubin to provide unbiased estimates of the variances of coefficients. In the examples below we set \( Q = 5 \), which should be sufficient to provide stable variance estimates.

The Adventist Health Study is a large cohort study that aims to enroll more than 100,000 subjects. As the major hypotheses are dietary, missing data are a significant problem, even when each food has only a small proportion of missing values. This is problematic because not only are analytic models often highly multivariate, but individual variables (eg, nutrient indices) will usually be the weighted sum of information from many foods. The food frequency questionnaire contains approximately 130 hard-coded food items and allows up to 50 write-ins. This questionnaire was based on extensive pilot work in this population. Each item has a given standard portion size, but allows subjects to nominate a larger portion (≥1.5 standard) or a smaller portion (≤0.5 standard). There are 7–9 frequency categories for each food.

As a test situation, we identified a sequential sample of the first 24,000 questionnaires and a set of key variables that include the 4 most influential foods in each of 18 indices for vitamins, nutrients, and minerals. These indices had been constructed in a pilot database (320 subjects), using cross-validation techniques to select the influential foods.

Within 6 months of receipt of the questionnaires, we attempted to contact a random 20% of all subjects who had one or more of the key variables missing. This contact was by telephone. Such contacts were possible for 90–95% of the study members, depending on the variable. The resulting filled-in data are used in the third illustrative model described below.

The first linear regression analysis has log(BMI) as the dependent variable and the following independent variables: meat (4 categories), nuts (5 categories), exercise (5 categories), age and sex. “Meat” and “nuts” are composite variables, each the sum of several questions about meats and nuts. The frequency data from each of these component questions were summed using midpoint values of categories, and then the composite variable was categorized.

The second model has the same dependent variable, but with the following independent variables: energy-adjusted folate (ea-folate), energy-adjusted polyunsaturated fatty acids (ea-PUFA), exercise (5 categories), age, and sex. Energy-adjustment used the residual method, and missing residuals were imputed.

To demonstrate that the method works, 2 quite artificial situations were set up (Tables 1 and 2), which did, however, start with real data. There were 20,126 subjects who had no missing data for the variables in these 2 models. This is the reference population for a complete data analysis and the parent population in subsequent analyses.

The imputation software required that continuous variables be normally distributed. However, log(BMI) was not normal, with skewness of 0.50 and standardized kurtosis.
of 0.47 (both equal 0 in a normal distribution). Thus, a new set of normally distributed body mass index (BMI) values was substituted. These were calculated from the regression E(log[BMI]) independent variables in the model) + e. The expected values were based on regression coefficients, where e’ is the error and e ~ N(0, var(e’)). Next, an artificial missing data mechanism that is decidedly not missing at random is imposed. For the first analysis (Table 1), all 2823 subjects in the higher meat categories 3 and 4 who had BMI values greater than or equal to 27 had their meat data converted to missing (15% of the total). Thus the association between BMI and meat in remaining subjects with observed values of 3 and 4 for meat will poorly predict meat consumption in those with missing data.

In Table 2, we use similar methods for the second model where the 2 energy-adjusted nutrients ea-folate and ea-PUFA, rather than a food, have missing data. Again, the data were transformed so that log(BMI), ea-folate and ea-PUFA are multivariate normal conditional on the covariates age, sex and exercise. Values of error variances and covariances observed in the non-normal data were used as parameters in the multivariate normal distribution.
Artificial missing data mechanisms are imposed on \(ea\text-\)folate and \(ea\text-\)PUFA, resulting in 21\% missing for \(ea\text-\)folate and 24\% missing for \(ea\text-\)PUFA. These were

1. for \(ea\text-\)folate, \(pr(missing|ea\text-\)folate \(> \) mean (\(ea\text-\)folate), BMI \(\geq 27) = 1, otherwise 0;
2. for \(ea\text-\)PUFA, \(pr(missing|ea\text-\)PUFA \(> \) mean (\(ea\text-\)PUFA)) = K, otherwise 0; where K = \(-0.21 + 0.0071 \times \) age, the coefficient chosen so K = 0 at age 30 and 0.50 at age 70 years.

The mechanisms are not missing at random because for both variables \(pr(missing)\) depends on the value of data that may be missing, specifically the requirement that missing values be greater than the mean for that variable.

Finally, in a third model we use the actually observed and naturally missing data from all 2752 black subjects within this database. Indicators of missing data are constructed for consumption of nuts and meat. The frequency of originally missing data for nut variables is 12\%, and for meat variables is 9\%. Log(BMI) is the dependent variable of the regression. The very few subjects with missing sex, age, or BMI were excluded. About 95\% of the 20\% sample of initially missing data has since been filled in, and was used to guide the imputation.

We performed the imputation using software available free on the internet,\(^ {12}\) which uses the Monte Carlo Markov Chain method\(^ {16}\) to estimate parameters of the likelihood function. Specifically, Schafer’s MIXED routine was used for this imputation, as BMI and age are continuous but some other variables are categorical.

### RESULTS

Table 1 shows the beta coefficients and their standard errors for the reference model, where BMI is transformed to normality. A second analysis simply excludes the missing data, a third imputes all missing meat data assuming missing at random (patently incorrect here), and finally the imputations are guided by subsample data. In Table 2, similar analyses are shown for the second model, where missing data for \(ea\text-\)folate and \(ea\text-\)PUFA are also decidedly not missing at random.

These results show that, as expected, simply excluding cases with missing data, or assuming that data were missing at random when they were not, produced beta coefficients for meat, and to a lesser extent, other variables, that were very biased. However, when the random new data were used to guide the imputation, even with these rather extreme missingness mechanisms, the point estimates were almost identical to the reference results. This was true not only for the meat, \(ea\text-\)folate and \(ea\text-\)PUFA coefficients, but also for other variables with no missing data whose coefficients were sometimes distorted if subjects with missing data were simply excluded. Such a case is the coefficient for exercise 3 in model 1, which, when missing data are excluded, changes by nearly 2 standard deviations. As expected, the standard errors for meat3, meat4, \(ea\text-\)folate and \(ea\text-\)PUFA are a little wider with imputation than the reference.

Sensitivity of these results to moderate non-normality in a variable assumed to be normal (such as BMI) is shown when using the non-normal log(BMI) as dependent variable. For variables not needing imputation, there is still good agreement between the reference results and those after imputation guided by the random sample. However, the new reference coefficients (with SE) for meat 2, meat 3, and meat 4 are 0.043 (0.0035), 0.076 (0.0037), and 0.102 (0.0049), whereas after imputation the results are 0.043 (0.0035), 0.082 (0.0041), 0.090 (0.0061). Thus, the meat3 and meat4 coefficients are nearly 2 standard errors removed from the reference results. Repeating the analysis several times with a different random 20\% sample consistently produces similar differences. Correcting the non-normality removed these discrepancies (Table 1).

Similarly, in the second model when non-normalized data are used for log(BMI), \(ea\text-\)folate, \(ea\text-\)PUFA, the new reference coefficients for the last 2 variables are \(-0.032 (0.0044)\) and \(0.251 (0.095)\), but after imputation, using the 20\% filled-in missing data, the coefficients are \(-0.039 (0.0053)\).
and 0.209 (0.100). These also differ moderately from the reference data, presumably due to the non-normality, as normalized data do not show these discrepancies (Table 2).

The results of Table 3 are those where the naturally missing data were imputed using the random 20% of filled-in missing data in the real subsample. There are modest changes in the beta coefficients when a missing at random assumption is made, as against the reference result where the new filled-in data are used. These changes are often 5–10% of the reference values. Imputing zeroes produced moderate changes for meat, but only small changes in the nut variables. Excluding cases with missing data produced greater changes in the nut variables but only small changes in meat variables. Standard errors were usually higher in this last case as there was less nonmissing data retained in the data set.

In these analyses, we somewhat arbitrarily chose to fill in 20% of missing data. We now investigate the effects of filling in only 10% or 5% instead (Table 4). This change should not affect validity, but decreases precision. However, as the filled-in data affects only imputed values for the remaining missing subset, effects on precision will also depend on the proportions of originally-missing data. If there is very little missing data, the standard error of the coefficients will change little regardless of the proportion of missing data filled in.

We show the effects of varying the proportion of data filled in for the artificial-missing situations (the models of Tables 1 and 2) where 15–24% of data was missing, as well as for the natural missing situation of Table 3, where 8–12% of data was originally missing. Where the proportion of missing data is greater (Models 1 and 2), filling in only 5% of missing data adversely affected statistical power, increasing standard errors by about 50% for some variables. However, where the proportion missing is smaller (Model 3), the smaller changes in standard errors suggest that filling a lower proportion of the missing data (a less costly choice) would be appropriate.

### DISCUSSION

When a random sample of originally missing data is filled in and the imputation model properly adjusted, the results of a reference analysis with no missing data can be accurately restored by guided multiple imputation. This result is of course expected from statistical theory. Even so, multiple imputation has rarely been used in epidemiology, despite the pervasive problem of missing data. Instead, most studies exclude subjects with much missing data, assume that missing data in food frequency questionnaires represent zeroes, or impute mean values from those not having missing data.17–19

While these procedures may perform satisfactorily20 in some situations, there can be no assurance of that without gathering further data. In other situations, such strategies are likely to be suboptimal. It is probable that more accurate imputation especially for frequently-eaten foods may efficiently improve the validity of many nutrient indices.

Information from several nutritional data sets4,18,20 suggest that, on average, about half the true values of initially missing nutritional data are not zero. This is particularly so for more common foods,4,18 and their missing data may not always be represented well by imputing the mean or median from the data that were initially observed. Further, the data necessary to evaluate the adequacy of a particular strategy are usually exactly that which we have used to give the more accurate imputation.

When different analytic approaches are applied to a real dataset with naturally-missing data (Table 3), results show less distortion than with the contrived situations of Tables 1 and 2. This is due both to smaller proportions of missing data, and to data that are probably closer to missing at random. Even so, the biases observed are of moderate magnitude in some instances, and probably worth some effort to avoid.

Glynn et al4,9 demonstrate the success of multiple imputation when incorporating information from a secondary nonrespondent survey. They used data only from this survey to impute the missing data, thus mandating different scaling coefficients for the nonrespondents; only the outcome variable was assumed to have missing values. While no doubt appropriate in their application, we argue that the coefficients used for imputation do not need to differ between the “nonrespondent” and “respondent” groups in the context of an epidemiologic study. All are actually respondents who happen to omit a usually small proportion of data. If the analyst prefers to make different imputations for the 2 groups, the imputation model can be adjusted to allow product terms between the predictor variables and the missing-ness indicator. The need for these can be evaluated using the subsample data. In the data of the analysis in Table 3, there was no convincing evidence that beta coefficients differed between data ini-
tially present and that filled in later. The new data can also be used to evaluate the adequacy of an assumption of missing at random, for data that were initially missing.

In the context we describe, all subjects in the subsample have filled out a long questionnaire at least moderately well and intend to continue to participate in the study. Indeed, in our cohort study it has proven realistic to fill in about 95% of the originally missing data from a random sample of 20% of those with at least one missing key variable. Thus, the ideal of complete responses from such a random sample can be approached quite closely. This approach is in contrast to the use of a secondary sample of nonrespondents to add more subjects to a survey, perhaps using shorter surveys or incentives to prompt a response.21

Ideally, the proposed application should be planned as an initial strategy to deal with missing data. If great efforts have already been made to fill missing data, completing a random sample of those still remaining will fall short by far, as these by definition are all very difficult-to-contact or poor responders. Usually, the resources to contact all subjects with missing data in a large cohort are not available. Hence, it is preferable to contact a random sample of these as an initial strategy. (Note that the method is not appropriate for missing data in 24-hour recalls or food diaries in which the missing data are specific to a time period and cannot be supplied retrospectively by a subsample.)

This more satisfying method of dealing with missing data is at some cost, as even filling a random sample of missing data will take some effort. Indeed it will usually be impossible to completely fill such data, because during the interval between original data collection and recontact some people will die, a few may develop serious illness, and others will decide not to cooperate with the subsample contact. Thus, it is acknowledged that the method will not allow extrapolation with confidence to those cohort members who correspond to (hopefully few) nonresponding subsample subjects.

How complete does the response in the random sample of missing data need to be? The proportional bias (Appendix) in the mean of subsample R when representing this by an incomplete subset, is P(1 − P)[1 − E(R')/E(R)], where E(R') is the mean of those in the subsample those data cannot be filled in, and P is the corresponding proportion. Thus, as expected, this proportional bias depends on P, and also on how unusual the missing portion is, reflected here by the ratio E(R')/E(R). Most would probably be satisfied with filling in 90% or more. For instance, if P = 0.1, and even if E(R')/E(R) is as improbably small as 0.5, the proportionate bias in the subsample is only 0.055. In fact, the imputation may also depend heavily on data not initially missing.

The size of the random sample R, for which missing data should be filled in, will depend in part on the required precision. A variable with only a small proportion of data initially missing will cause little loss in precision when using the multiple imputation procedure. For variables with more missing data it will be wise to ensure that a sufficient proportion is filled in (Table 4). Filling a tiny random percentage will produce unbiased results, but the standard errors will be wide. Calculation of the statistics r and lambda proposed by Schaefer7 revealed similar trends. A more quantitative but somewhat heuristic approach to this question is found in the Appendix.

One could cautiously conclude that, when less than 10% are missing, it is important to impute to avoid losing data, but within reason the form of the imputation may not make a lot of difference to the outcome. Other results18 are consistent with this approach. Our experience in Adventist Health Study-1,22 and more recently in Adventist Health Study-2, as well as the experience of others,23 is that 3–10% of values typically are missing for particular food frequency items in an older population. The frequent use of compound indices in nutritional epidemiology exacerbates this situation. It is when the proportion of missing values is moderate or large that the effects of different epidemiologic applications on bias and precision are more influential.

Multiple imputation can handle large proportions of missing data where necessary, but assuming that data are missing at random when they are not would then be particularly hazardous. The application described here will be helpful when the missing at random assumption is not clearly warranted, and where there is at least a moderate amount of missing data.

In summary, where the proportion of missing data is modest (typical of many epidemiologic applications), and where it is not clearly missing at random, filling in a random sample of 10% of the missing may be sufficient to guide the imputation. In our experience this process does not create a large budgetary burden, and allows satisfactory bias correction without appreciable loss of power. If normality is assumed, there is some sensitivity to deviation from this assumption. Appropriate transformations, such as the Box-Cox,24 can be used. Alternatively, such variables could be converted to categorical forms.

REFERENCES

APPENDIX

Proportional Bias in Subsample Data Created by Incomplete Response

For variable X, let subscripts c and ic indicate successfully completed and finally incomplete subsample data; n indicates number of subsample subjects, and P is the proportion of subsample subjects whose data is not filled-in due to nonresponse.

Then \( \frac{[E(X_c) - E(X)]/E(X)} \) is the proportionate bias considering only the completed subsample data.

\[
\frac{(n_c + n_{ic})E(X) - n_cE(X_c)}{n_cE(X)} - 1
\]

\[
= \frac{n_{ic}(E(X) - E(X_c))}{n_cE(X)}
\]

Dividing numerator and denominator by \( n \) this becomes

\[
P \left( \frac{1 - E(X_c)}{1 - E(X)} \right)
\]

Why Both the Proportion of Initially Missing Data, and the Proportion of This That is Filled in Affects the Precision of the Final Analysis

The number of nonmissing data points is one measure of the information available to the analysis. For large numbers, \( Var(\theta) \) is approximately proportional to \( \frac{1}{(N - D(1 - L)) - m} \), where \( D \) is the proportion initially missing, \( L \) is the proportion of these filled in later, and there are \( m \) variables in \( \theta \). Hence, \( L = R/DN \). Obviously \( Var(\theta) \) smaller as \( L \) increases or as \( D \) becomes smaller. A smaller \( Var(\theta) \) will result in Q imputation data sets that are increasingly similar. Consequently \( Var[E(\beta)] \), over the Q runs, is smaller for a given \( D \), where \( \beta \) is the coefficient vector of the analysis performed after the imputation. As \( Var[E(\beta)] \) is the quantity by which the missing data increases the variance of \( \beta \) when using multiple imputation, \( Var(\beta) \) is smaller when \( L \) is larger or \( D \) is smaller. A related observation is that with either of these trends there will also be fewer data points to impute. As these carry the influence of \( Var(\theta) \), the corresponding imprecision can have less impact. Another implication is that when \( D \) is large, choosing a larger \( L \) will reduce the resulting imprecision.
Alternative Methods of Estimating an Incubation Distribution

Examples From Severe Acute Respiratory Syndrome

Benjamin J. Cowling,* Matthew P. Muller,‡‡ Irene O. L. Wong,* Lai-Ming Ho,* Marie Louie,‡ Allison McGeer,‡‡ and Gabriel M. Leung*

Background: Accurate and precise estimates of the incubation distribution of novel, emerging infectious diseases are vital to inform public health policy and to parameterize mathematical models.

Methods: We discuss and compare different methods of estimating the incubation distribution allowing for interval censoring of exposures, using data from the severe acute respiratory syndrome (SARS) epidemic in 2003 as an example.

Results: Combining data on unselected samples of 149 and 168 patients with defined exposure intervals from Toronto and Hong Kong, respectively, we estimated the mean and variance of the incubation distribution allowing for interval censoring of exposure intervals between times A and B, where A ≤ X ≤ B, is usually documented. If symptom onset occurred at time Z, interest lies in the incubation time T = Z-X, but we can only observe that symptoms began between L = Z-B and R = Z-A days after infection.

These data are a special type of survival data, and a natural approach would be to “reverse” the time axis setting Z as the origin and X as the outcome time. Then each patient’s incubation time is not exactly known, but must be in the interval (L, R). As De Gruttola and Lagakos explained, “reversing” the time axis is “valid only when the density function for infection is uniform in chronicologic time.” This condition is reasonable in the setting of SARS, with each exposure interval being relatively short, and will be assumed to hold for all the analyses described below; in fact, this assumption underlay all previous estimates of SARS incubation either implicitly6–13 or explicitly.14 Further details on this assumption are given in the technical appendix at the end of this article.

In the following sections we describe nonparametric and parametric statistical techniques for accurate estimation of the incubation distribution and illustrate the bias in some commonly used methods. We present the first inferential

During the 2003 severe acute respiratory syndrome (SARS) epidemic, knowledge of the incubation period of SARS was critical to the global effort to control the spread of illness. Based on initial estimates of incubation time, a 10-day quarantine period was adopted by most health authorities, although some suggested that a longer quarantine period might have been more appropriate.1 Furthermore, the incubation period was incorporated into the case definition for SARS early in the outbreak: patients were only considered to have SARS if they had a potential exposure within the 10-day period before symptom onset.2 Estimates of the incubation distribution were also required for developing the mathematical models used to understand transmission dynamics and to predict the impact of different control policies.3 Thus, accurate estimates of the incubation distribution, and particularly the right-hand tail in the first 2 scenarios, are central in informing public health policy.

Estimation of the incubation period is complicated because infection events cannot be directly observed. Instead, patients typically report a period of exposure during which infection likely occurred; if the infection occurred at time X, an exposure period between times A and B, where A ≤ X ≤ B, is usually documented. If symptom onset occurred at time Z, interest lies in the incubation time T = Z-X, but we can only observe that symptoms began between L = Z-B and R = Z-A days after infection.
METHODS

The simplest estimate of the incubation distribution is obtained by excluding cases in which exposure occurred within a defined interval, and using data only from cases in which a single exposure occurred at a defined time. However, this approach may underestimate the incubation distribution because of recall bias, since recent exposures are likely to be more precisely remembered. To incorporate defined exposure periods, the infection dates could be imputed as the midpoint of those exposure intervals, which then permits empirical estimation. Alternatively, one could assume that all infections occurred at the beginning or end of each interval, and then calculate the corresponding empirical distribution functions.

Turnbull derived the nonparametric maximum likelihood estimator of the distribution function for interval-censored exposure data, which simplifies to the empirical distribution function if all exposure times are exactly observed. As the maximum likelihood estimate, this approach should be viewed as the nonparametric gold-standard. It can be implemented in R (Development Core Team, Vienna) through the package “icen,” or in SAS (SAS Institute Inc., Cary, NC). Pointwise 95% confidence intervals may be calculated from the observed information matrix. An alternative simple nonparametric estimate of the incubation distribution for interval-censored data is given by Meltzer.

It is often useful to summarize the incubation distribution by a parametric model, which can easily accommodate interval-censored exposure data, and which simplifies to the empirical distribution function if all exposure times are exactly observed. As the maximum likelihood estimate, this approach should be viewed as the nonparametric gold-standard. It can be implemented in R (Development Core Team, Vienna) through the package “icen,” or in SAS (SAS Institute Inc., Cary, NC). Pointwise 95% confidence intervals may be calculated from the observed information matrix. An alternative simple nonparametric estimate of the incubation distribution for interval-censored data is given by Meltzer.

A further advantage of parametric models is that covariate information can be incorporated through the scale parameter, and these models are known as accelerated life or accelerated failure time models. When the lognormal distribution is used, the corresponding regression model is essentially a multiple linear regression, or equivalently an analysis of covariance, on the natural logarithm of the incubation time. Given that incubation distributions are typically right-skewed and always positive-valued, it is more appropriate to perform linear regression on log-transformed incubation times than on untransformed times. The estimated regression coefficients, , can be interpreted as the expected change in median log incubation time relative to baseline, while the transformed effects exp() can be interpreted as acceleration factors ie, proportional increases or decreases in the median incubation time. Parametric regression models for interval-censored data may be applied in R (through the package “survival”) or in SAS (using “proc lifereg”). All analyses presented here were conducted using R version 2.3.1 (R Development Core Team, Vienna).

APPLICATION TO SARS DATA

Data on a defined exposure period from Toronto were available from 149 (51%) of all 291 adults diagnosed with “probable SARS,” whereas similarly specified data were available from 168 (10%) of the 1755 probable SARS patients in Hong Kong.

Figure 1A compares the Turnbull nonparametric estimate and 3 parametric models on the Toronto data. By Akaike’s Information Criterion, the best-fitting distribution was the lognormal (AIC = 478), followed by the Weibull (AIC = 480), and then the gamma (AIC = 482). Figures 1B and 1C show the same distributions fitted to the Hong Kong data and the combined data respectively, where again the lognormal fitted best. Visual inspection of the parametric curves against the Turnbull estimates in Figures 1A–C confirm that the lognormal distribution fits the data appropriately. For the combined data, the Turnbull estimate had mean and variance 5.0 days and 11.5 days, respectively, and the 95th percentile was 11.0 days. Using the lognormal parametric model for the combined data, the estimated mean and variance were 5.1 days and 18.3 days, respectively, and the 95th percentile was 12.9 days (95% confidence interval = 11.7–14.5 estimated by bootstrapping).

The estimated means, 95th and 99th percentiles of the lognormal, Weibull, and gamma distributions are presented in Table 1, with 95% confidence intervals generated by bootstrapping. The 95th and 99th percentiles would be of interest for public health planning, in particular for specifying case definitions and the quarantine period. All 3 models have fairly similar values of AIC, but quite significant discrepancies in the tails, where the fitted lognormal distributions have the longest right-hand tails. We also investigated the log-gamma distribution which is a generalization of the gamma and lognormal distributions, and found the results were almost identical to the lognormal model in each case (data not shown). As a sensitivity analysis, we fitted lognormal models to the incubation data for only the 128 and 136 laboratory-confirmed SARS cases in Toronto and Hong Kong and found that the estimates were very similar to those for all probable SARS patients (Figs. 2A and B). As a further sensitivity analysis, we fitted a lognormal model to data on exactly known exposures only (n = 65) and found that the mean and variance of the SARS incubation period were 5.1 days (95% CI = 4.2–5.9) and 16.5 days, respectively, and the 95th percentile was 12.6 days (CI = 10.1–15.4).

Previous estimates of the incubation distribution are summarized in Table 2 and compared with the estimates for data presented here. Early estimates from Hong Kong, and estimates from Singapore provided the shortest incubation times whereas a recent estimate from Beijing is consistent with our results.
TABLE 1. Alternative Parametric Estimates of the Mean and Right-Hand Tail of the Incubation Distribution in Hong Kong (n = 168), Toronto (n = 149), and Combining All Data

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Mean (Days)</th>
<th>95th Percentile (Days)</th>
<th>99th Percentile (Days)</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>(95% CI*)</td>
<td>Estimate</td>
<td>(95% CI*)</td>
</tr>
<tr>
<td>Combined</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lognormal</td>
<td>5.1</td>
<td>(4.6–5.5)</td>
<td>12.9</td>
<td>(11.7–14.5)</td>
</tr>
<tr>
<td>Weibull</td>
<td>5.0</td>
<td>(4.6–5.4)</td>
<td>11.4</td>
<td>(10.4–12.5)</td>
</tr>
<tr>
<td>Gamma</td>
<td>4.2</td>
<td>(3.8–4.6)</td>
<td>10.4</td>
<td>(9.6–11.3)</td>
</tr>
<tr>
<td>Lognormal</td>
<td>4.4</td>
<td>(3.7–5.1)</td>
<td>12.4</td>
<td>(10.2–14.9)</td>
</tr>
<tr>
<td>Weibull</td>
<td>4.3</td>
<td>(3.7–5.0)</td>
<td>11.2</td>
<td>(9.5–13.0)</td>
</tr>
<tr>
<td>Gamma</td>
<td>3.2</td>
<td>(2.6–3.7)</td>
<td>9.6</td>
<td>(8.3–11.1)</td>
</tr>
<tr>
<td>Toronto</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lognormal</td>
<td>5.6</td>
<td>(5.0–6.2)</td>
<td>12.3</td>
<td>(10.8–14.0)</td>
</tr>
<tr>
<td>Weibull</td>
<td>5.5</td>
<td>(5.0–6.1)</td>
<td>11.3</td>
<td>(10.1–12.5)</td>
</tr>
<tr>
<td>Gamma</td>
<td>5.0</td>
<td>(4.5–5.5)</td>
<td>10.7</td>
<td>(9.6–11.7)</td>
</tr>
</tbody>
</table>

*95% CIs calculated by bootstrapping with 5000 repetitions.
AIC indicates Akaike Information Criterion.
Figure 1D compares the alternative nonparametric estimates of the incubation distribution. Imputing the infection date as the midpoint of exposure intervals overestimated the median of the incubation distribution compared with Turnbull’s estimate, and the bias was worse under Meltzer’s nonparametric estimate. Finally, the 2 extreme empirical estimates, where infection dates were imputed as either the left or the right end of all exposure intervals, gave slightly wider bounds for the incubation distribution than the 95% pointwise confidence intervals of the Turnbull estimate. Figure 1E compares the Turnbull estimates of the estimated incubation distributions in Hong Kong and Toronto, while the lognormal estimates are compared in Figure 1F.

We performed a multiple linear regression on the log of the incubation times, allowing for interval censoring. The resulting regression coefficients and acceleration factors are shown in Table 3. Older patients had 45% longer median incubation times, while patients in Hong Kong and healthcare workers had 21% and 34% shorter median incubation times than patients in Toronto and nonhealthcare workers, respectively. These findings were fairly consistent under separate analyses for each center (Appendix Table).

**TABLE 2.** Estimates of the SARS Incubation Distribution From Different Countries by Different Methods

<table>
<thead>
<tr>
<th>Location (Reference)</th>
<th>Patients Analyzed</th>
<th>Method</th>
<th>Incubation Distribution (Days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hong Kong6</td>
<td>57</td>
<td>Parametric (gamma)</td>
<td>Mean: 3.8; 95th Percentile: 9.5</td>
</tr>
<tr>
<td>Singapore8</td>
<td>198</td>
<td>Parametric (Weibull)</td>
<td>Mean: 4.8; 95th Percentile: 9.7</td>
</tr>
<tr>
<td>Singapore7</td>
<td>94</td>
<td>Empirical with midpoint assumption</td>
<td>Mean: 5; 95th Percentile: 9.7</td>
</tr>
<tr>
<td>Toronto8</td>
<td>59</td>
<td>Empirical excluding interval-censored</td>
<td>Mean: 4.7; 95th Percentile: ≤10*</td>
</tr>
<tr>
<td>Toronto9</td>
<td>42</td>
<td>Empirical excluding interval-censored</td>
<td>Mean: 5; 95th Percentile: ≤10†</td>
</tr>
<tr>
<td>Hong Kong10</td>
<td>128</td>
<td>Parametric (loggamma)</td>
<td>Mean: Not given; 95th Percentile: 10.7</td>
</tr>
<tr>
<td>Canada, Hong Kong, and U.S.11</td>
<td>19</td>
<td>Nonparametric (Meltzer)</td>
<td>Mean: ~5; 95th Percentile: 12</td>
</tr>
<tr>
<td>Hong Kong12</td>
<td>81</td>
<td>Parametric (gamma)</td>
<td>Mean: 4.6; 95th Percentile: 12.5</td>
</tr>
<tr>
<td>Beijing13</td>
<td>209</td>
<td>Empirical with midpoint assumption</td>
<td>Mean: 5.3; 95th Percentile: 13.9</td>
</tr>
<tr>
<td>Toronto, Hong Kong14</td>
<td>317</td>
<td>Parametric (lognormal)</td>
<td>Mean: 5.1; 95th Percentile: 12.9</td>
</tr>
<tr>
<td>Toronto15</td>
<td>149</td>
<td>Parametric (lognormal)</td>
<td>Mean: 5.6; 95th Percentile: 12.4</td>
</tr>
<tr>
<td>Hong Kong16</td>
<td>168</td>
<td>Parametric (lognormal)</td>
<td>Mean: 4.4; 95th Percentile: 12.5</td>
</tr>
</tbody>
</table>

*The 95th percentile was not given, but only one patient (1.7%) had an incubation time greater than 10 d.
†The 95th percentile was not given; the range of possible incubation times was 2–10 d.
‡Current study.

**DISCUSSION**

We have estimated the incubation distribution of SARS based on 317 patients with known exposure periods, which is the largest combined database in the literature to date. The estimated mean and 95th percentile are consistent with other recent estimates in the SARS literature, while slightly longer than the earliest published estimates (Table 2). Two previous empirical analyses of small subsets of SARS patients in Toronto with exact infection times found mean incubation times of 5 and 4.7 days,8,9 slightly shorter than the 5.6 day...
mean in the full Toronto data analyzed here. Previous analyses of the incubation data in Hong Kong typically excluded longer exposure intervals, and this may have biased those estimates towards shorter times if recent exposures were more precisely recalled than earlier exposures. It is also noticeable from Figure 1 and Table 1 that the gamma distribution used in those previous estimates has a shorter right tail than the lognormal and Weibull distributions, and thus a shorter 95th percentile. When only the exact exposure times were analyzed, we found that the estimated incubation distribution was very similar while the smaller sample size resulted in a less precise estimate of the 95th percentile.

We found that the 95th percentile of the incubation distribution was 11 days under the nonparametric estimate, but 12.9 days under the best-fitting parametric model. However, the means of the nonparametric and parametric estimates were very close. A potential reason for the shorter right tail of the nonparametric estimate could be because the World Health Organization clinical case definitions for SARS required exposure within 10 days, and therefore some earlier World Health Organization clinical case definitions for SARS right tail of the nonparametric estimate could be because the mates were very close. A potential reason for the smaller distribution was very similar while the smaller sample size resulted in a less precise estimate of the 95th percentile. When only the exact exposure times were analyzed, we found that the estimated incubation distribution was very similar while the smaller sample size resulted in a less precise estimate of the 95th percentile.

We found that the 95th percentile of the incubation distribution was 11 days under the nonparametric estimate, but 12.9 days under the best-fitting parametric model. However, the means of the nonparametric and parametric estimates were very close. A potential reason for the shorter right tail of the nonparametric estimate could be because the World Health Organization clinical case definitions for SARS required exposure within 10 days, and therefore some earlier potential exposures (ie, longer incubation times) may not have been reported, or even if reported these may not have been recorded in the charts due to the perceived implausibility of long incubation time. Furthermore, the lack of exposure data for many patients, particularly in Hong Kong, could have biased the estimates of the incubation distribution. We found that the lognormal distribution gave best fit to the SARS data, and this is biologically plausible since this distribution has given best fit to the incubation distribution of a wide variety of infectious diseases. We investigated the incubation distribution using data on probable SARS cases to be consistent with the literature, and we note that our results were very similar when we analyzed the subgroup with laboratory confirmation of SARS.

The alternative parametric models varied quite significantly in the estimated length of the right-hand tail of the incubation distribution (Table 1). The 95th percentile, or a similar statistic from the right-hand tail of the incubation distribution, is important for public health planning for 2 reasons. First, it would allow proper specification of the case definitions and minimize misdiagnosis of patients with long intervals between exposure and onset of symptoms of the disease in question. Second, it would allow proper specification of the quarantine period, essentially the length of time that an exposed but asymptomatic person will be inconvenience before they can be assumed to be disease free. Given that public health officials should aim for conservatism, a useful approach might be to fit a variety of parametric models, and not only select the best-fitting model (according to robust criteria) but also look at the longest (ie, most conservative) tail percentiles.

We found significant geographical variation in the incubation distribution between Hong Kong and Toronto, even after adjusting for other important factors (Table 3). Analysis of incubation data from China also found significant geographical variation and hypothesized that these may have been due to differences in host resistance, or differences in socio-demographic or environmental characteristics in the various areas. An ex ante alternative explanation is that geographical variation in the incubation period could have been due to variation in genotypic differences of the SARS virus although subsequent investigation showed that the Metropole Hotel strain from Hong Kong seeded the worldwide outbreak, including Toronto. Another explanation for the differences between Hong Kong and Toronto is that reporting behavior was different, either due to differences in recognition of initial symptoms or differences in elicitation of the exposure data. Finally, as commented previously, the patients who provided exposure data may not have formed a representative sample, which could have led to biases in the estimated incubation distributions in either or both locations.

Our results suggest that incubation times varied with age and occupation. The longer incubation periods experienced by older patients might have resulted from a delayed immune response (albeit the disease consequences were more severe with age), given that SARS has characteristics of an immunomodulated disease. The shorter incubation periods of healthcare workers might have been due to the exposure of healthcare workers to patients with high viral loads, as peak viral load and peak infectivity typically occurred at day 7 to day 10 of illness, a time point at which most SARS patients had already been hospitalized. Alternatively, these findings may be due to reporting bias if younger individuals or healthcare workers were more vigilant for symptom onset. While an earlier analysis of incubation data in China did not find any statistically significant relationship between these variables, the direction of effects in their results is consistent with those presented in Table 3.

Turnbull’s nonparametric maximum likelihood estimate should be viewed as the gold standard estimate of the incubation time.

### TABLE 3. Factors Affecting the Log Incubation Time of Severe Acute Respiratory Syndrome in Hong Kong and Toronto (n = 317)

<table>
<thead>
<tr>
<th>Factor</th>
<th>Acceleration Factor*, Exp(β)</th>
<th>Estimate (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yrs)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤39†</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>40–59</td>
<td>0.94</td>
<td>(0.77–1.15)</td>
</tr>
<tr>
<td>60+</td>
<td>1.45</td>
<td>(1.15–1.83)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female†</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.99</td>
<td>(0.83–1.19)</td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not healthcare worker†</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Healthcare worker</td>
<td>0.79</td>
<td>(0.65–0.96)</td>
</tr>
<tr>
<td>Location</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toronto†</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Hong Kong</td>
<td>0.66</td>
<td>(0.55–0.79)</td>
</tr>
</tbody>
</table>

*The acceleration factor indicates the proportional increase (>1) or decrease (<1) in the median time from infection to onset of symptoms.

†Reference category.
incubation distribution and used in preference to midpoint imputation whenever possible. For incubation distributions that are right-skewed, as is typically the case,\textsuperscript{19,20} the midpoint imputation approach will overestimate the incubation distribution. However, midpoint imputation should provide practical estimates during the early stages of an emerging epidemic when data are scarce and any estimate will have considerable uncertainty, noting that overestimation should be preferable to underestimation. As a final caveat, estimation of the incubation period can be further complicated if the onset date is unclear although for SARS this was not a problem given that the defining symptom was fever.\textsuperscript{22}

In emerging infectious diseases, accurate and precise estimates of the distribution of incubation times are necessary to advise public health policy, to specify case definitions, and to facilitate robust mathematical modeling. In this paper we have described and illustrated the alternative methods which could be used to derive the incubation distribution of SARS. The statistical techniques discussed here are generalizable to any infectious disease where exposure is not exactly known and thus incubation data are interval censored.

ACKNOWLEDGMENTS

The authors thank their colleagues in the Hong Kong Department of Health and Hospital Authority and in Toronto who were involved with the public health control of the SARS epidemic and data collection and processing.

REFERENCES


APPENDIX

A Discussion of the Assumption of Uniform Probability of Infection

If infection occurred on the interval (A, B), and symptoms occurred at Z, then the likelihood contribution for an individual is of the form

\[ \frac{B}{u=A} g(u) f(Z-u) \, du, \]

where \( g(u) \) describes the probability of infection at calendar time \( u \), and \( f(\cdot) \) is the density of the incubation distribution. This simplifies to the expression \( \{F(Z-A) - F(Z-B)\} \), where \( F(\cdot) \) is the cumulative density of the incubation distribution, if \( g(u) \) is the uniform density.\textsuperscript{5} Therefore the assumption of constant infection probability through any given exposure interval allows us to “reverse the time axis,” since the simplified expression above is identical to the likelihood contribution for survival data which are censored on the interval (Z-B, Z-A).\textsuperscript{16}

De Gruttola and Lagakos\textsuperscript{5} propose a method for joint estimation of \( g(u) \) and \( f(\cdot) \), where \( g(u) \) represents the changing population risk of infection; however, this does not seem appropriate in the example of SARS, in which population risk was negligible throughout the epidemic. Individual risk may have varied considerably (eg, when healthcare workers were on or off duty) but such detailed data on precise exposure intervals are rarely available. As an example of a situation where the population risk of infection \( g(u) \) is of interest, De Gruttola and Lagakos\textsuperscript{5} estimate the increase in risk of HIV over calendar time for hemophiliacs using contaminated blood banks.

© 2007 Lippincott Williams & Wilkins
The sensitivity of the estimated incubation distribution to the assumption of constant risk in the 2 most extreme cases of decreasing (increasing) risk can be shown by calculating the empirical distribution functions assuming that all infections occurred at the beginning (end) of individual exposure intervals. The extreme estimates for the SARS example are shown in Figure 1D, and are somewhat wider than the 95% pointwise confidence intervals about the Turnbull estimate.

Fitting Parametric Models to Interval Censored Survival Data

We have considered 3 parametric models for the incubation distribution:

1. The lognormal distribution with parameters $\mu$, $\sigma$ and probability density function

   $$f(x) = \exp\left(-\frac{(\ln(x) - \mu)^2}{2(\sigma^2)}\right).$$

2. The Weibull distribution with parameters $r$, $\lambda$ and probability density function

   $$f(x) = x^{r-1} \exp\left(-x/\lambda\right).$$

3. The gamma distribution with parameters $k$, $\theta$ and probability density function

   $$f(x) = x^{k-1} \exp\left(-x/\theta\right) / \Gamma(k)/\theta^k.$$

Each of these distributions may be fit to data which include incubation times written in the form $[L_i, R_i]$ for individuals $i = 1, \ldots, n$, where for $L_i < R_i$ these are interval censored times and for $L_i = R_i$ they are exactly observed incubation times. Note the special conventions in notation $L_i$ for left-censored incubation times and $R_i$ for right-censored incubation times. Then each of the parametric models may be fit by maximizing (eg, by numerical optimization) the likelihood function of the form

$$L(\lambda, \theta \mid \text{data}) = \prod_{i : L_i < R_i} (F(R_i) - F(L_i)) \prod_{i : L_i = R_i} f(L_i),$$

where $F(\cdot)$ and $f(\cdot)$ are the relevant cumulative distribution function and probability density function, respectively.4

| APPENDIX TABLE. Estimated Acceleration Factors for SARS, by Age, Sex and Occupation, in Hong Kong (n = 168) and Toronto (n = 149) |
|---|---|---|---|
| Age (years) | Hong Kong Acceleration Factor* | Toronto Acceleration Factor* |
| | Estimate | (95% CI) | Estimate | (95% CI) |
| 39† | 1.00 | | 1.00 | |
| 40–59 | 0.72 (0.51–1.04) | 1.11 (0.88–1.40) |
| 60+ | 1.74 (1.19–2.55) | 1.21 (0.93–1.59) |
| Sex | | | |
| Female† | 1.00 | 0.97 (0.79–1.19) |
| Male | 1.00 (0.74–1.36) | |
| Occupation | | | |
| Not healthcare worker† | 1.00 | | |
| Healthcare worker | 0.78 (0.54–1.13) | 0.80 (0.65–0.99) |

*The acceleration factor is computed as $\exp(\beta)$. It indicates the relative increase ($>1$) or decrease ($<1$) in the median time from infection to onset of symptoms.

†Reference category.
Modeling Infectious Diseases Dissemination Through Online Role-Playing Games

Ran D. Balicer

Abstract: As mathematical modeling of infectious diseases becomes increasingly important for developing public health policies, a novel platform for such studies might be considered. Millions of people worldwide play interactive online role-playing games, forming complex and rich networks among their virtual characters. An unexpected outbreak of an infectious communicable disease (unplanned by the game creators) recently occurred in this virtual world. This outbreak holds surprising similarities to real-world epidemics. It is possible that these virtual environments could serve as a platform for studying the dissemination of infectious diseases, and as a testing ground for novel interventions to control emerging communicable diseases. (Epidemiology 2007;18: 260–261)

Mathematical modeling is important in the study of infectious diseases and the development of public health policies. Even so, agent-based modeling and similar simulations are limited in their potential to account for changes in human behaviors during epidemics. This has led to searches for novel methods to simulate human daily interactions.1,2

One possibility lies in existing Internet role-playing games. The numbers of participants in these virtual worlds have grown rapidly since 1996, involving millions of players engaged in interactive play. Each world develops its own economy, with production, assets, and interactive trading with Earth economies.3 These virtual worlds are realized in meticulous detail, and are inhabited by characters played by real people worldwide, as well as by nonplayer characters controlled by set programs. These different types of characters have varied and complex options for interacting with each other—an environment that may be suitable for testing dissemination patterns of infectious diseases.

The possibility of epidemics in these virtual worlds is not merely academic. A serious epidemic of an infectious disease4 recently erupted among the virtual characters in “World of Warcraft,” one of the most popular multiplayer online games.5 Although the disease had been created by game administrators, its consequences were unanticipated.4 Information on this outbreak was retrieved from players’ forums, allowing insights on dissemination patterns in this virtual outbreak and on the potential of these virtual environments for modeling infectious disease outbreaks and control.

OUTBREAK DETAIL

The outbreak began on late September 2005 when the “World of Warcraft” game administrators introduced a new virtual creature that had the ability to cast a disease (“corrupted blood”) on its opponents. In addition to inflicting severe damage on the target character, the disease “infected” close contacts who could spread the disease to others in close proximity.5 Game administrators presumably believed that the short period of infectivity (several seconds), as well as its highly lethal effect, would render the disease self-limiting. This proved not to be the case.

The pandemic plague that resulted is unique. Unlike previous “virtual plagues” that had been officially planned, this was a local effect that went out of control—a naturally occurring virtual outbreak. In the words of one player, “What happened next was just plain weird. When infected adventurers returned to town at the end of their quest, they inadvertently passed along the Corrupted Blood infection to those nearby. In short order, the plague ravaged the population. Game administrators were baffled. As they scrambled to quarantine areas of the game world, the disease quickly spread beyond their control. Partially to blame was the game’s feature that allows players to teleport from one area to another, and which made it possible for the plague to rapidly reach the most distant regions of the map.”6

According to information from various Internet blogs, several epidemiologic attributes enabled this uncontrolled dissemination of the disease. One was the lack of residual immunity following convalescence. This enabled characters to be reinfected and re-enter the transmission cycle. The second characteristic was its infectivity to the virtual animals (“pets”). While pets were relatively resistant to the lethal effects of the disease, they were infective to other pets and humans, thus serving as a reservoir. Continuous cycles of the disease between pets and humans could therefore allow the infection to simmer until the group reached densely populated areas. Third, ill characters could teleport, thus introducing a disease with short infectivity period through large distances. Lastly, once the plague reached the cities, it did not just infect other players but also the nonplayer characters of the city, providing a large “bystander” population that also spread the disease.
COMMENT

This plague is the first unplanned large-scale virtual infectious disease in game worlds. There have been smaller unplanned outbreaks. In May 2000 players of “The Sims” were outraged when their game characters died of an infection contracted from virtual guinea pigs that had not been adequately cared for. However, this recent outbreak was substantially larger and more complex in its evolution and outcomes. This could be in part due to the detailed environment and character interactions in “World of Warcraft.”

Some observers have suggested that the unexpected spread of this virtual infection was the deliberate strategy of malicious players. It is possible that players were able to sustain the transmission cycle of the disease by keeping in close contact with another player while constantly healing each other until they reached populated cities. If so, this incident may also count as the first virtual act of bio-warfare.

Mathematical modeling experts may find some utility in this outbreak as a case study, applying models to predict future disease dissemination and adequacy of various interventions. The basic reproductive rate in virtual illnesses can be quantified, the population-specific force of infection can be assessed, and predicted outcomes can be compared with actual events influenced by player-dependent behaviors.

Multiplayer online role-playing games may even be useful as a testing ground for hypotheses about infectious disease dissemination. Game programmers could allow characters to be infected by various infectious diseases, some of which may not be visible to the player, and track the dissemination patterns of the disease in specific subpopulations. For example, modes of transmission could be defined for automated nonplaying characters whose behavior and mixing patterns are set, as well as for human-directed characters whose behavior is difficult to predict. Various interventions could be tested, including treatment (either freely dispensed or sold), vaccination, isolation and quarantine.

While the parallels with a real-world outbreak are striking, the artificial nature of the games limits them as models for the real world and might even lead to misleading conclusions about real infectious outbreaks. The mixing patterns and interactions among the game figures may be considerably different from those in real life, and would depend heavily on the rules and goals of the game. The most obvious example would be the risk-taking behavior of virtual characters, which depends heavily on the penalties for death or illness and the availability of a “game saving” option.

**FIGURE 1.** Epidemic in the World of Warcraft.

Despite these limitations, there could be advantages to studies in virtual worlds. The mixing patterns and behavior observed in the game can be precisely measured and accounted for (without the usual epidemiologic problems of incomplete ascertainment or loss to follow-up). Furthermore, the rules and environment could potentially be adjusted to allow better modeling of specific real-life scenarios. (Admittedly, this approach depends on the ability to make these changes without undermining the essential pleasures of the game.) Expert modelers of infectious diseases might consider collaborating with the game’s administrators. Such collaborations could harness the immense computational power invested in these economically-driven, large-scale virtual environments, while allowing simulations more wide-ranging than any options currently available.

The game’s administrators eventually cured the plague with a “spell” that was distributed rapidly to players en masse. If only real life were that simple.

**REFERENCES**

The Fate of Epidemiologic Manuscripts
A Study of Papers Submitted to *Epidemiology*

Susan A. Hall* and Allen J. Wilcox†

**Background:** Little is known about the success rate of epidemiologic manuscripts, or the number of rejections they may go through before being published.

**Methods:** In late 2004 we conducted a retrospective follow-up study of the cohort of manuscripts submitted to *Epidemiology* in 2002. Using an e-mailed invitation, we conducted an online survey of authors identified from journal records. Authors were asked about submission attempts before and after their submission to *Epidemiology*.

**Results:** *Epidemiology* received 371 original articles in 2002, of which it published 101 (27%). Survey response rates were 68% among the authors of accepted manuscripts, and 58% among authors of manuscripts rejected by *Epidemiology*. These responses provided a total sample of 223 manuscripts for analysis. Of the cohort, 83% (n = 184) were eventually accepted for publication (by *Epidemiology* or others). The acceptance rate by *Epidemiology* was the same whether or not the manuscripts had been previously rejected by another journal. Of the 155 manuscripts rejected by *Epidemiology*, 116 (75%) were eventually published or accepted for publication, 11 (7%) were being prepared for resubmission at the time of follow-up (19–34 months after rejection), 5 (3%) were under review by a journal, and 23 (15%) were inactive. Among the papers we could follow from first submission, 62% of those eventually published had been rejected at least once. In general, papers rejected by one journal were subsequently sent to a journal with lower impact factor.

**Conclusions:** These data suggest most epidemiology manuscripts are eventually published, although some persistence on the part of the authors may be necessary.

(*Epidemiology* 2007;18: 262–265)

The opportunities for successful publication of scientific manuscripts would seem to be abundant, given the proliferation and expansion of scientific journals. However, very few longitudinal cohort studies of research manuscripts have been conducted, and there are few data available on the actual publication rate of scientific papers. Such studies are difficult because there is no practical way to enumerate a cohort of initially submitted manuscripts.

Likewise, little is known about the path that successful manuscripts take on the road to publication. A few studies have described selected cohorts of manuscripts that were identified during the review process. These studies suggest that the majority of manuscripts are eventually published, although success rates vary.1, 4 We use data from a survey of authors who submitted manuscripts to *Epidemiology* to reconstruct the journey and final status of this cohort of manuscripts.

**METHODS**

Our study population comprised all authors who submitted an original scientific manuscript to *Epidemiology* in calendar year 2002. Names and e-mail addresses of corresponding authors were identified using journal records. We contacted authors via e-mail in late October and early November 2004, providing them with an explanation of the survey’s purpose, the title of their submitted manuscript, and a direct link to a web-based questionnaire (set up through Zoomerang/ZPro5). We chose a web-based questionnaire for its convenience, and also because we thought its novelty would appeal to this research-savvy group of participants. Survey questions are provided in an online supplement to this paper, available on the journal’s web site. In a pilot survey among epidemiologists, the questionnaire required 5–10 minutes to complete.

In the survey, we asked participants whether their manuscript had been submitted to other journals before its submission to *Epidemiology* and, if so, to which journals. Authors whose manuscripts were declined by *Epidemiology* were asked additional questions about their manuscripts (if any) to which their manuscripts were subsequently submitted, and the current status of their manuscripts. Authors were invited to contact us with any questions or concerns. We sent one reminder e-mail notice to nonresponders.

An explicit question to obtain informed consent was built into the survey tool. The study was approved by the Institutional Review Board of the School of Public Health of the University of North Carolina at Chapel Hill (September 2004). In addition to the survey data, we used impact-factor data from Thomson Scientific as of 2004. We conducted simple descriptive analyses, which included the calculation of mean and median impact factors for groups of journals.

From the *Department of Epidemiology, University of North Carolina at Chapel Hill, and †Epidemiology, Durham, North Carolina. Submitted 3 November 2006; accepted 14 November 2006. Susan A. Hall’s epidemiology postdoctoral fellowship was funded by National Cancer Institute Grant Number 2 T32 CA009330-26. Correspondence: Susan A. Hall, New England Research Institutes, 9 Galen Street, Watertown, MA 02472. E-mail: sahall@email.unc.edu Copyright © 2007 by Lippincott Williams & Wilkins ISSN: 1044-3983/07/1802-0262 DOI: 10.1097/01.ede.0000254668.63378.32

Epidemiology • Volume 18, Number 2, March 2007

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.
RESULTS

EPIDEMIOLOGY received 371 unsolicited original scientific manuscripts between 1 January and 31 December 2002 (not including letters, commentaries, book reviews, and other material). Of these, 101 were accepted for publication in EPIDEMIOLOGY (27%). Our survey response rate was 68% (n = 69) among the corresponding authors of the 101 accepted papers, and 58% (n = 157) among the authors of 270 rejected manuscripts. Of these respondents, we had to exclude 3 who later told us their response was in regard to a manuscript other than the one named. This left a total of 223 manuscripts for analysis. The follow-up period for manuscripts declined by EPIDEMIOLOGY (i.e., time from rejection to survey interview) ranged from 19 to 34 months.

Of the 223 manuscripts with response data, we had complete submission history for 220. (We assigned the minimum number of possible rejections to the 3 manuscripts with incomplete history; these have trivial effects on the subsequent analyses.)

A total of 137 manuscripts (62%) had no history of submissions before their submission to EPIDEMIOLOGY. The remaining 83 manuscripts had been previously rejected by at least one journal. Within the total cohort of 223, the acceptance rate at EPIDEMIOLOGY was 30% (n = 68). (This acceptance rate is slightly higher than the acceptance rate for all submissions to EPIDEMIOLOGY in 2002 [27%], owing to the higher response rate among authors whose manuscript had been accepted.) The acceptance rate was the same for manuscripts previously rejected by another journal (30%) and for manuscripts being submitted for the first time (31%).

The subsequent history of manuscripts rejected by EPIDEMIOLOGY is shown in the flow chart (Fig. 1). Of the 155 manuscripts rejected, 140 (90%) were subsequently resubmitted or being prepared for resubmission elsewhere at the time of follow-up. The remaining 15 (10%) were never resubmitted (“inactive”). After a minimum of 19 months follow-up, 116 of the rejected manuscripts (75%) had been published or accepted for publication, 11 (7%) were being prepared for resubmission, while a total of 23 (15%) had become “inactive.” Taking the cohort of 223 manuscripts as a whole, 83% (n = 184) were eventually accepted for publication (by EPIDEMIOLOGY or by another journal), 7% were still being worked on, and 10% (n = 23) were inactive.

We also considered the subcohort of 137 manuscripts that had no previous history of submission before EPIDEMIOLOGY. Although we accepted only 31% (n = 131) were eventually accepted within the period of follow-up. (Another 6% [n = 8] were still in process, and 12% [n = 16] had become inactive.) Among the 113 published papers in this subcohort, 70 (62%) were rejected by at least one journal (EPIDEMIOLOGY), and 11% were rejected by at least 2 journals before being published.

Journals and Their Impact Factors

Among manuscripts submitted to more than one journal, there was a general pattern of progression from higher- to lower-impact journals. In 2004, EPIDEMIOLOGY had an impact factor of 4.2. The 27 journals that rejected manuscripts before their submission to EPIDEMIOLOGY had a mean impact factor of 8.3 (median 4.5, range = 0.9 to 35). (Eleven percent of authors did not provide the title of the rejecting journal. Manuscripts declined by EPIDEMIOLOGY were subsequently submitted to 95 journals with a mean impact factor of 2.5 (median 2.3, range = 0.3 to 7.2). (19% of the authors reporting a subsequent rejection after EPIDEMIOLOGY did not provide the journal title, while only 1 of the 116 with a successful subsequent submission did not provide the publishing journal’s name.)

The AMERICAN JOURNAL OF EPIDEMIOLOGY (AJE) was the journal most likely to see a manuscript either before or after it had been submitted to EPIDEMIOLOGY. Of 26 manuscripts that had been previously rejected by AJE, EPIDEMIOLOGY accepted 6; in turn, AJE accepted 5 of 9 manuscripts that had previously been rejected by EPIDEMIOLOGY. The other journals most likely to review a manuscript before EPIDEMIOLOGY were THE LANCET (n = 12) and the BRITISH MEDICAL JOURNAL (n = 7). Other journals likely to receive a manuscript after EPIDEMIOLOGY were ANNALS OF EPIDEMIOLOGY (n = 6) and the EUROPEAN JOURNAL OF EPIDEMIOLOGY (n = 6).

DISCUSSION

Our cohort of manuscripts is a selected sample in several respects. Obviously, as a cohort of manuscripts submitted to EPIDEMIOLOGY, it represents a narrow spectrum of all biomedical manuscripts, and an incomplete sample even of all epidemiologic manuscripts. The effects of this selection cannot be addressed within our data.
Furthermore, the observed cohort does not represent all manuscripts that might have been submitted to Epidemiology, but rather only those that were not accepted elsewhere before being submitted to Epidemiology. Our cohort of manuscripts has thus been filtered by the priority authors gave to Epidemiology among all the journals that they regarded as appropriate and by the manuscript’s success at any higher priority journals. (Of the manuscripts submitted elsewhere before being submitted to Epidemiology, we can observe only those that were rejected. This amounts to informative left-censoring of our cohort.)

A complete description of the fate of a cohort of manuscripts would require the identification of a cohort of manuscripts from their creation—that is, from the time of their preparation for first submission. Such a study is unlikely ever to be done. Unfortunately, all other designs (including ours) raise questions about selection bias. Information on rejection status and access to authors resides with journal editors, and previous studies have been conducted by or in conjunction with journal staff. Ray et al studied a sample of manuscripts rejected by Annals of Internal Medicine in 1993 and 1994 and found that 69% were eventually published, with a trend towards more specialized journals and decreasing impact factor. Opthof et al examined manuscripts submitted to Cardiovacular Research in 1995 and 1996, and found 70% of original manuscripts were rejected. OVID was subsequently searched for titles of rejected manuscripts, and 47% were found to be published in other journals, with a lower mean impact factor. Finally, Nemery considered manuscripts rejected by Occupational and Environmental Medicine from 1995 to 1997. Of the 44% of manuscripts that were rejected over these years, 54% were traced later in MEDLINE, with 90% of those published papers appearing in a journal with a lower impact factor.

In our study, we relied on author report of publication, and we did not validate author report with Medline tracing. As has been previously noted, not all journals are contained within MEDLINE, so that MEDLINE searches may underestimate the subsequent acceptance rate.\(^1,3,4\) Furthermore, manuscripts may evolve in their author order or title in ways that make the paper difficult to identify through searches.

We are unaware of specific descriptions of success rates for epidemiologic manuscripts, other than an editorial in AJE that compared acceptance rates with Epidemiology. Considering one volume of AJE (June–December 2004), the editors reported receiving 446 original contributions and accepting 15% of these.\(^2\)

Perhaps the most striking limitation of our own study is the poor response rate (58% among the authors of manuscripts rejected by Epidemiology, and 68% among authors of accepted manuscripts). Other studies have reported low response rates for web surveys of professionals. A web survey of British general practitioners reported a response rate of 52% following 5 e-mail reminders.\(^6\) A study of response rates among surgeons reported a lower response rate in the web-survey arm (45%) compared with the mailed questionnaire arm (58%).\(^7\) Similarly, in studies conducted among staff of U.S. federal statistical agencies, poorer overall response rates were found for e-mailed surveys (43%) versus mailed paper surveys (70%).\(^8\) Although convenient to use, our e-mail survey may have appeared less legitimate than a paper survey on journal letterhead. Also, 3 authors reporting having technical difficulties with the survey.

The lower response rate among the authors of rejected manuscripts is particularly problematic. Authors who ultimately failed to publish their manuscripts may have been particularly unmotivated to participate in our survey. To the degree that such “failed” manuscripts are underrepresented in our cohort, our estimates of ultimate success rates will be inflated. We explored this possible bias by assuming that the entire difference in response rates between the 2 groups of authors (58% versus 68%) was due to authors whose rejected manuscripts were never published (an additional 10% from among the 270 rejected manuscripts). If we add these hypothetical 27 nonresponders with failed manuscripts to our cohort of 223, the total proportion published would be 74% (184/250) instead of 83%, and the total inactive would be 20% (50/250) instead of 10%, with another 6% still in process. Similarly, among the hypothetical cohort of 182 manuscripts (155 + 27 inferred) rejected by Epidemiology, 64% were ultimately published and 27% were inactive.

The generalizability of our data is limited in other ways as well. Our cohort is a mix of manuscripts previously submitted to (and rejected by) other journals and manuscripts being submitted for the first time. As discussed above, the former group is selected in that we cannot observe papers that would have been submitted to Epidemiology had the manuscript not been accepted elsewhere. Interestingly, the proportion of papers eventually achieving publication did not differ for these 2 groups (83% and 82%). However, even the subgroup of manuscripts being submitted for the first time cannot be regarded as representative of all epidemiologic manuscripts, since this group includes only those for which our journal was the author’s first choice. Author’s preference is probably not unbiased with regard to the author’s previous experience with journals, and the topic and quality of the paper, any of which might be related to ultimate success of the paper.

While our results provide incomplete and no doubt biased estimates for the unknown cohort of all epidemiologic manuscripts, these data do suggest certain patterns. First, as has been observed in previous studies, authors of a rejected manuscript on average turn to journals with lower impact factors. This may be because authors presume such journals have higher acceptance rates (although we know of no data regarding that correlation). This trend was not universal: some manuscripts rejected by Epidemiology were subsequently published in journals with higher impact factor, and some papers rejected by a journal with a lower impact factor were subsequently accepted by Epidemiology. Our impact factor analysis was limited in that authors were less able to recall of the name of journals that rejected their papers than those that accepted their papers.

More importantly, it appears that authors who continue to submit their manuscripts to journals after rejection have a reasonable chance of publication. The direction of cause-and-effect is not clear; successful publication may be the reward
of persistent authors, or better papers may provide the authors with confidence to continue to resubmit. Also, helpful reviewers may allow the manuscript to be improved along its path to acceptance.

While no amount of persistence can be expected to compensate for a fatally flawed paper, authors should be reassured by the fact that most published papers have a history of prior rejection. Among the papers we could follow from their first submission, 62% of those eventually published had been rejected at least once. If researchers are condemned to publish or perish, they should be encouraged to know that one or 2 rejections are not a death sentence.

ACKNOWLEDGMENTS

The authors thank Teresa P. Edwards of the Odum Institute for Research in Social Science for her assistance with this research.

REFERENCES


IN THE NEXT ISSUE

Neurobehavioral development in children with presumed exposure to pesticides

Work schedule during pregnancy and spontaneous abortion

Vitamin B₁₂ and the risk of neural tube defects

Effect of soy isoflavones on endometriosis: Interaction with estrogen receptor 2 gene polymorphism

Occupational exposures and breast cancer among Shanghai textile workers

Cooked meat and risk of breast cancer: Lifetime versus recent dietary intake

Confined animal feeding operations and respiratory health of neighboring residents

Short-term effects of temperature on cardiovascular deaths in the US elderly: Changes over time

Risk factors for positive tuberculin skin test in Guinea-Bissau

Studying vulnerable populations: lessons from the Roma minorities

Properties of Two Different Counterfactual Effect Definitions for a Point Exposure

Identification of synergism in the sufficient-component cause framework

Abe and Yak: Abraham Lilienfeld and Jacob Yerushalmy and the Development of Modern Epidemiology (1945-1973)
Background: Exposure to 60-Hz magnetic fields may increase breast cancer risk by suppressing the nocturnal production of melatonin. The use of medications associated with reduced melatonin levels could modify this relationship.

Methods: We recontacted participants in a population-based case–control study of residential magnetic field exposure and breast cancer risk and interviewed them regarding medication use during the 10 years before diagnosis. Cases were diagnosed between November 1992 and March 1995, and magnetic field levels were measured in the home at diagnosis. We obtained medication use information by telephone interview from 558 cases and 588 controls.

Results: Breast cancer risk was not associated with exposure to residential magnetic fields, regardless of medication use.

Conclusions: These results support previous findings that magnetic field exposure does not increase breast cancer risk.

It has been proposed that chronic exposure to 60-Hz magnetic fields may increase the risk of breast cancer by suppressing the normal nocturnal production of melatonin; this suppression may affect the development of breast cancer either by increasing levels of circulating estrogen or through an oncostatic property of melatonin itself. In a previous population-based case–control study of breast cancer (denoted the “original study”), we found no relationship between several measures of residential magnetic field exposure and breast cancer risk. However, a companion study of a subset of the controls found decreased nocturnal urinary melatonin levels associated with residential magnetic field exposure, which was particularly evident among women who were current users of calcium channel blockers, beta blockers, and psychotherapeutics. Furthermore, medication use was independently associated with reduced melatonin levels.

These results suggested that there may be variability in individual susceptibility to the effects of magnetic field exposure in lowering melatonin concentration. If so, medication use could modify an association between breast cancer risk and magnetic field exposure. Thus, we recontacted participants from the original study to inquire about medication use during the 10 years before diagnosis (cases) or corresponding reference date (controls) so that we could investigate whether breast cancer risk was increased with exposure to magnetic fields among women who use these medications.

METHODS

The original study, described more fully elsewhere, was a population-based case–control study of breast cancer that included 813 women 20–74 years of age with incident breast cancer who were diagnosed between November 1992 and March 1995 (cases) and 793 controls who were frequency-matched to the cases by 5-year age groups. The primary purpose of the study was to investigate whether exposure to residential magnetic fields or light-at-night increased breast cancer risk. At the time of diagnosis, magnetic field measurements were made in the participant’s bedroom using an EMDEX II meter (Enertech Consultants, Campbell, CA) over a 48-hour period. An in-person interview was used to ascertain information about known or suspected risk factors for breast cancer. Data collection for the original study occurred from April 1993 through December 1995.

Use of the aforementioned medications, as well as corticosteroids and nonsteroidal anti-inflammatory drugs, was ascertained via a telephone interview administered from March 2000 through December 2001. To disguise the primary focus of interest, questions also were asked about the use of vitamin supplements and about exercise habits. To facilitate recall, a copy of the questionnaire was provided before the scheduled telephone interview so participants could note names, dates, and frequencies of use for each medication. Participants were instructed to report only regular use, which was defined as use at least 4 days/wk for 6 months or longer. Each medication class was asked about individually, and common medical conditions were listed as examples for prescribing each type. If the participant reported regular use defined in this manner, she was then asked to list each medication and the beginning and ending dates of use.

Participants were instructed to report only regular use, which was defined as use at least 4 days/wk for 6 months or longer. Each medication class was asked about individually, and common medical conditions were listed as examples for prescribing each type. If the participant reported regular use defined in this manner, she was then asked to list each medication and the beginning and ending dates of use.

The Fred Hutchinson Cancer Research Center Institutional Review Board approved the procedures for recontacting study participants and all data collection.

As described previously, a variable was constructed to reflect mean magnetic field level in the bedroom during the night (defined for all participants as the 7 hours between 10 PM and 5 AM the following morning). Analyses regarding magnetic field measurements were restricted to nighttime to...
correspond to the underlying biologic hypothesis that exposure to magnetic fields may increase breast cancer risk by reducing the normal nocturnal secretion of melatonin.

Ever-use and duration-of-use variables were constructed for each medication class. Calcium channel blockers and beta blockers were analyzed as one medication type. Psychotherapeutics included antidepressants, anti-anxiety, antianxiety, anxiolytic, antipsychotic, and psychostimulants. Medication use indicators were used as stratification variables to investigate whether a potential relationship between breast cancer risk and nighttime bedroom magnetic field level differs according to medication use. Odds ratios and 95% confidence intervals were used to estimate relative risks using conditional logistic regression (SAS procedure PHREG, SAS/STAT Release 8.2; SAS Institute, Inc., Cary, NC). All models were conditional on 5-year age strata, with adjustment for factors associated with breast cancer risk previously identified in this study. Odds ratios were evaluated using the Wald \( \chi^2 \) test.

**RESULTS**

Of the 813 cases and 793 controls in the original study, bedroom magnetic field measurements were obtained for 744 cases and 711 controls (92% and 90%, respectively). Of these, approximately 75% (n = 558) of the cases and 83% (n = 588) of the controls who were still living at the time of recontact completed the telephone interview (Table 1). Among all participants, use of nonsteroidal anti-inflammatory drugs was the most prevalent (approximately 24% of both cases and controls reported regular use); calcium channel blockers and beta blockers were the second most prevalent medication used (15% of cases and 13% of controls), followed closely by psychotherapeutics (13% of cases and 12% of controls). Relatively few participants reported corticosteroid use (4% of cases and 5% of controls).

There was no evidence of an association between increasing nighttime bedroom magnetic field level and breast cancer risk, regardless of ever-use or duration of use of any of the medications of interest (Table 2). We also attempted to investigate whether women who used one of these medications and were exposed to magnetic fields at higher levels (0.3 and 0.4 \( \mu \) T) were at an increased risk for breast cancer. Although there was no evidence of increased risk in those most highly exposed, the number of participants in this analysis was very small.

**DISCUSSION**

This study found no evidence of increased breast cancer risk associated with increasing nighttime bedroom magnetic field level, regardless of ever-use or duration of use of several classes of medications previously shown to be associated with reduced melatonin levels. These findings are in agreement with our previous study of breast cancer and residential magnetic field exposure, as well as others.

Previously, we found that the reduction in nocturnal melatonin levels associated with increasing bedroom magnetic field level was more pronounced among users of the medications investigated in this study. We also found that the strongest magnetic field effects were observed during the summer months when melatonin levels were lowest, regardless of medication use. Furthermore, the magnetic field effect increased somewhat with older age (which was associated with lower melatonin levels). Collectively, these findings suggested that persons with lower baseline melatonin levels—for whatever reason—may be more susceptible to the additional effects of magnetic fields in lowering their melatonin concentration and thus perhaps are more susceptible to the effects of magnetic field exposure on breast cancer risk. Consistent with these findings are results from an experiment in which reductions in melatonin associated with exposure to 20-\( \mu \) T magnetic fields were observed only in those with low baseline melatonin levels, (although 2 subsequent experiments failed to replicate this finding), as well as results from a study that found considerable individual variability in sensitivity to the effects of light-at-night on melatonin levels in humans.

Three methodologic limitations of the present study could potentially account, at least in part, for the absence of an association between breast cancer risk and magnetic field exposure among medication users. First, because participants were recontacted approximately 5 to 8 years after the original study and asked to report medication use during the 10 years before diagnosis, they had to recall medication use 5 to 18 years earlier. This undoubtedly resulted in some degree of misclassification regarding categories of medication use, which is likely to bias the risk estimates towards the null. However, to facilitate recall, participants were provided an advanced copy of the questionnaire and were asked to note details of use before the telephone call. Furthermore, a recent study, also conducted in Washington State, compared patient self-reported antihypertensive medication use to pharmacy records and concluded that self-reported use was relatively accurate among older women, specificity among drug classifications was high, and recall did not differ by case–control status.

Second, there were relatively few participants who reported regular use of the medications of interest, and thus...
small to modest increases in risk among the medication users would be difficult to detect in this study. Third, our investigation of the effect on breast cancer risk of a possible interaction between magnetic field exposure and medication use is indirect. We did not measure nocturnal melatonin levels and therefore cannot evaluate directly whether such levels are associated with medication use, magnetic field exposure, or breast cancer risk (although one recent study did find higher melatonin levels associated with reduced breast cancer risk20).

In summary, these findings, based on data collected from a large population-based study, add to the considerable body of evidence that exposure to magnetic fields is not associated with breast cancer risk, even within specific subgroups of women who might be more susceptible to the effects of such exposure.

ACKNOWLEDGMENTS

The authors thank the following individuals for their valuable contributions to this work: Laurie Shields, project...
REFERENCES

Background: Occupational and experimental animal studies indicate that exposure to high levels of manganese impairs male fertility, but the effects of ambient manganese in humans are not known.

Methods: We measured blood levels of manganese and selenium in 200 infertility clinic clients in a cross-sectional study. Correlations between metals and semen variables were determined, adjusting for other risk factors. Outcomes were low motility (<50% motile), low concentration (<20 million/mL), or low morphology (<4% normal). We also investigated dose–response relationships between quartiles of manganese exposure and sperm parameters.

Results: High manganese level was associated with increased risk of low sperm motility (odds ratio = 5.4; 95% confidence interval = 1.6–17.6) and low sperm concentration (2.4; 1.2–4.9). We saw a U-shaped dose–response pattern between quartiles of manganese exposure and all 3 sperm parameters.

Conclusion: Ambient exposure to manganese levels is associated with a reduction in sperm motility and concentration. No adverse effects were seen for high selenium.

Manganese is a ubiquitous transition metal found naturally in the environment. It is also released into the air from mining and manufacturing operations and from combustion of gasoline additives. Human exposure to ambient levels of manganese is universal and occurs mainly via air and dust exposures. Although trace levels of manganese are required for normal sperm function, high levels have been shown to adversely affect male fertility. Occupational studies have found that chronic high manganese levels in male workers resulted in impotence, decreased birth rates, and decreased semen parameters. Animal and in vitro experiments indicate that high manganese exposure decreases sperm motility and concentration, possibly via membrane lipid peroxidation. To determine whether ambient levels of manganese are associated with measures of male infertility, we measured manganese levels in the blood of men attending infertility clinics in Michigan and assessed their relationships to sperm motility, concentration, and morphology. Although sperm motility is imperfect as a predictor of infertility, it has been shown to be a good discriminator between men of proven fertility and infertile patients with high sensitivity (0.74) and specificity (0.90).

METHODS

The subjects of this cross-sectional study were the first 200 participants in an ongoing study on environmental contaminants and measures of male infertility recruited at 2 infertility clinics in Michigan between 2003 and 2005. The male partner of every couple seeking infertility care was offered participation in this study; men were not selected based on whether they contributed to the couples’ infertility status. After obtaining informed consent, we enrolled men who were 18 to 55 years of age and currently attempting to conceive a pregnancy with their partner. Exclusion criteria for the entire project were diabetes, thyroid or adrenal disorders, genetic disorders related to fertility, a history of testicular cancer or bilateral orchiectomy, and hormonal therapy. The protocols for this study were approved by the research ethics committees at all participating institutions.

A questionnaire ascertained demographics, lifestyle, occupation, and general and reproductive health history. Men with occupations suggesting manganese exposure (n = 22) did not differ from the rest of the population in their metal levels or semen parameters. Participants provided semen and blood samples within 1 week of enrollment. Semen samples were collected and analyzed using the World Health Organization protocols. Three andrology technicians at 2 locations performed the semen analyses. Intra- and interlaboratory comparisons of sperm concentration and morphology showed no substantial differences, and sperm motility did not differ by site. Sperm motility was evaluated at 1 hour postcollection. Percent motile was the sum of the percentages with rapid linear progression (3–4+) and slow linear progression (2+). Morphology was assessed using Kruger’s strict criteria. Five azospermic men were excluded from the study.

Whole venous blood was drawn using stainless steel needles; collected into 2-ml glass tubes prescreened for mercury, cadmium, and lead; and stored at −20°C. Manganese and selenium standards were used to construct standard curves, and the curves were later used to extrapolate metal levels in the samples. Controls included human blood with...
low metal levels and bovine blood. Certified reference standards for manganese and selenium were not available. All samples were analyzed with an Elan DRC-plus ICP mass spectrometer (Perkin Elmer, Norwalk, CT), using the Centers for Disease Control and Prevention methodology. The limits of detection for blood manganese and selenium were 1.0 and 5 μg/L, respectively.

Sperm motility and subject age were normally distributed. We classified manganese and selenium into dichotomous variables for high levels, as defined by the 75th percentile (14.0 μg/L for manganese and 200 μg/L for selenium). Spearman correlation coefficients were determined for age, semen parameters, and metal levels. We compared mean motility, concentration, and morphology in men with high blood manganese versus all others using the Wilcoxon test and, in a second analysis, to men whose sperm concentration, motility, and morphology were all above the World Health Organization cutoff levels. Results of these analyses and reports in the literature were used to evaluate variables for inclusion in logistic regression models with low motility (<50% motile sperm),9 low concentration (<20 million sperm/mL),9 or low morphology (<4% normal forms).9 Variables were removed from the saturated model using backward selection, based on their negative impact on model fit. The relationship of quartiles of manganese levels to each of the sperm parameters was also assessed.

RESULTS

Mean ± SD sperm motility was 53% ± 15% (median = 55%), ranging from 0 to 85%. Mean ± SD sperm concentration was 56 ± 48 million/mL (median = 41), ranging from 0.3 to 324 million/mL. Sperm morphology had a mean ± SD of 4.0 ± 3.7% normal forms (median = 3.7%), ranging from 0 to 16%. Mean blood manganese was 12.2 μg/L (median = 12.5), ranging from 5.0 to 30.0 μg/L. Mean blood selenium was 187 μg/L (median = 184), ranging from 130 to 260 μg/L. Men with high manganese levels were less likely to be African American or to have smoked in the past 12 months, and were more likely to have a higher income (Table 1).

Age (−0.15) and manganese (−0.10) were negatively correlated with motility, and selenium (0.06) was positively correlated. Manganese (−0.08) was negatively correlated with concentration, and selenium (0.05) was positively correlated. Manganese (−0.06) was negatively correlated with morphology and was positively correlated with selenium (0.07).

Men with high manganese levels were more likely to have low sperm motility, concentration, and morphology and higher selenium levels as compared with men who had lower manganese levels (Table 2). When the relationships between manganese quartiles and the 3 semen parameters were examined, a U-shaped relationship emerged, with men in both the highest and lowest manganese quartiles at increased risk for lower semen parameters (Figure 1). In regression models, men with high manganese levels were at increased risk of low motility and low concentration (Table 3) when compared with other men. When the comparisons were restricted to

<table>
<thead>
<tr>
<th>TABLE 1. Demographic Characteristics by Blood Manganese Category</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High Manganese</strong></td>
</tr>
<tr>
<td>Race; (%)</td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td>African American</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>Income category; (median)</td>
</tr>
<tr>
<td>Age; (mean ± SD)</td>
</tr>
<tr>
<td>Smoking; (%)</td>
</tr>
<tr>
<td>Current</td>
</tr>
<tr>
<td>Last 12 months</td>
</tr>
</tbody>
</table>

*High manganese level was defined as at or greater than the 75th percentile of the manganese distribution for the study population.

<table>
<thead>
<tr>
<th>TABLE 2. Distribution of Low Semen Parameters and High Se Level by Blood Manganese Category</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High Manganese</strong></td>
</tr>
<tr>
<td>Sperm parameter</td>
</tr>
<tr>
<td>Low motility</td>
</tr>
<tr>
<td>Low concentration</td>
</tr>
<tr>
<td>Low morphology</td>
</tr>
<tr>
<td>High selenium</td>
</tr>
</tbody>
</table>

*High manganese level was defined as at or greater than the 75th percentile of the manganese distribution for the study population.

![FIGURE 1. Dose–response relationships between quartiles of blood manganese levels and the sperm parameters.](image)

© 2007 Lippincott Williams & Wilkins
participants with only low motility or only low concentration, compared with men who had normal values for all sperm parameters (n = 79), the OR for high manganese and sperm motility increased to 5.4 (95% CI = 1.6–17.6), but decreased to 1.3 (0.5–13.2) for low concentration. High manganese level was not associated with low sperm morphology. When each study site was analyzed separately, the ORs were similar and in the same direction.

**DISCUSSION**

Men with high (14 μg/L or higher) or low (<10 μg/L) manganese levels had low sperm motility and concentration but not low sperm morphology (Fig. 1). The dose–response analyses supported these observations and also showed that low manganese level was associated with low sperm motility and low concentration, although not as strongly. The low manganese effect is plausible in light of the critical role for manganese in many metabolic processes, including reproduction.2 The high manganese level is at or above the normal range for blood manganese (4–14 μg/L),1 whereas the low manganese level is within the normal range. Age was not a strong risk factor for decreased motility in our study, although other studies have shown inverse associations.16 It is possible that the relatively young age of the men in our sample (median age 34 years) may explain the lack of an age effect. Smoking is a risk factor for poor semen quality,11 but it was not associated with any sperm parameter in our analyses. Selenium had a protective effect, which is supported by its key role in spermatogenesis.12,13 These results suggest that ambient levels of manganese are a potential risk factor for low sperm parameters. The participation rate in this study was approximately 70%. It is unlikely that nonparticipants would differ systematically from participants in both manganese levels and semen parameters. Strengths of this study include the simultaneous analyses of several metals in 200 individual blood samples. Manganese adversely affected both sperm motility and concentration, and this effect was observed using several analytic methods, supporting the robustness of the relationship.

Several studies have detected manganese in human seminal fluid, but little information is available on the effects of ambient manganese exposure on human fertility or semen quality. In an occupational study, exposure to manganese salts significantly decreased the odds of having a live birth, but a second study did not find the same effect among workers exposed to high manganese levels in dust,3 possibly because of different blood manganese levels. Chinese mine workers occupationally exposed to manganese in dust (geometric mean 0.14 mg/m3) had semen with reduced liquefaction time, decreased sperm count, and decreased percentage of viable sperm, compared with men with the same occupations from the same area who were not exposed to manganese.1 An interactive effect of exposure to multiple metals, however, was possible.

Because of the heterogeneity of human sperm, it has been difficult to find a biomarker that can discriminate fertile men from fertile men. Sperm motility, however, has been found to more accurately identify men with male factor infertility than other semen parameters and is correlated with fertility status.8 The number of motile sperm has also been found to be highly predictive of intrauterine insemination success9 and motility is correlated to in vitro and in vivo fertilization rates. Motility was the semen parameter most tightly linked to high manganese levels in our study.

**REFERENCES**


Is Height Associated With Cardiovascular Risk in Chinese Adults?

C. Mary Schooling,* G. Neil Thomas,* Gabriel M. Leung,* Sai Yin Ho,* Edward D. Janus,† and Tai Hing Lam*

Background: Better childhood conditions, proxied by greater height, are usually protective against ischemic heart disease in western countries. These relations are less evident in other settings.

Methods: We used multivariable logistic regression to examine the relation of height to the metabolic syndrome and its components in a rapidly developed Asian population using a representative, cross-sectional Hong Kong Chinese sample of 2860 adults from 1994 to 1996.

Results: Height was inversely associated with increased blood pressure (odds ratio = 0.74; 95% confidence interval = 0.58–0.94) and raised fasting plasma glucose (0.71; 0.55–0.91), but only after adjustment for central obesity. Central obesity was also positively associated with height (2.09; 1.67–2.62) for tallest compared with shortest tertile, confounding these relationships. The association between height and central obesity was much stronger in men than in women.

Conclusion: The relation of height to cardiovascular risk may relate to a society’s history and stage of socioeconomic development.

In long-term industrialized populations, height is usually associated with lower mortality from ischemic heart disease (IHD),1,2 and lower morbidity from associated risk factors.3–5 The reasons for these associations potentially relate to improvements in childhood conditions. However, in more recently developed populations, the protective effect of height on IHD mortality6 or its risk factors7–10 is less obvious. Moreover, the social patterning of IHD in men in western populations reversed during the 20th century, from association with higher to lower social position.11 However, height has always been associated with social advantage,12 suggesting that the relationship between height and IHD may be ethnically or historically specific.

Hong Kong has a high standard of living and little social patterning of IHD.13 Economic development began in Hong Kong in the early 20th century.14 However, most Hong Kong residents are first- or second-generation immigrants from China, where the gross domestic product per capita was similar to preindustrialized western Europe until the second half of the 20th century.15 Thus, Chinese people in Hong Kong have experienced a recent epidemiologic transition compressed over far fewer generations than in long-term industrialized countries. Given the relatively recent history of economic development in Hong Kong (and China), there are no long-term cohort studies in which to examine height and IHD risk prospectively. Instead, to explore the etiologic role of height in IHD, we investigated the relationship between height and a composite marker of cardiovascular and IHD risk, ie, the metabolic syndrome,16,17 in Hong Kong Chinese.

METHODS

Materials

The sampling frame and methods of this population-based cross-sectional study in Hong Kong have been described elsewhere in detail.18 In brief, a probability sample was used to obtain the WHO MONICA19 recommendation of approximately 200 participants in each of 5 10-year, adult, sex-specific age groups. Randomly generated telephone numbers were used to select households, and within households, a Kish grid was used to randomly select respondents. Almost all of Hong Kong households (99%) have a telephone.20 From 1994 to 1996, 7174 Chinese (25–74 years) took part in telephone interviews (a response rate of 78%). The study was supplemented by a relatively small number (n = 556) of participants ages 55–74 who were recruited specifically because there were fewer people of this age in the population. Of the 7730 interviewed, 2900 (38%) came for physical examination. We excluded pregnant women and those with serious diseases such as cancer and those who were hospitalized. Those coming for physical examination quite closely matched the telephone sample and the population and bias should thus be small.18 For place of birth, exercise, smoking status, self-reported medically diagnosed diabetes, or hypertension and general health, the Cohen effect sizes were negligible (<0.1). Effect sizes for job activity, a proxy for socioeconomic status (0.16), and education (0.23) were slightly larger but still acceptable. The detailed methods of

Submitted 27 February 2006; accepted 27 October 2006.

From the *Department of Community Medicine and School of Public Health and †Clinical Biochemistry Unit, The University of Hong Kong, Hong Kong.

The present contact address for Edward D. Janus is Department of Medicine, Western Hospital, Footscray, VIC. 3011, Australia.

Supported by the Hong Kong Health Services Research Committee (HSRC# 411026), the University of Hong Kong Committee on Research and Conference Grants, the Hong Kong Research Grants Council (grant 407/94m), and the Hong Kong Society for the Aged.

Correspondence: G. Neil Thomas, Department of Community Medicine and School of Public Health, Faculty of Medicine Building, 21 Sassoon Road, Pokfulam, Hong Kong. E-mail: neilt@hkucc.hku.hk.

Copyright © 2007 by Lippincott Williams & Wilkins

ISSN: 1044-3983/07/1802-0274
DOI: 10.1097/01.ede.0000254656.02400.27
measurement have been reported. The University of Hong Kong Ethics Committee approved the study and participants gave written, informed consent.

Risk Factors
The metabolic syndrome was defined according to the International Diabetes Federation guidelines. Individuals were classified as having the metabolic syndrome if they had central obesity (waist circumference ≧80 for women or 90 cm for men) and at least 2 of the following: (a) raised blood pressure: systolic or diastolic blood pressures ≧130/85 mm Hg or treatment of previously diagnosed hypertension, (b) raised fasting plasma glucose of ≧5.6 mmol/L or treatment of previously diagnosed type 2 diabetes, (c) raised triglycerides: fasting plasma triglycerides ≧1.7 mmol/L or specific treatment, (d) reduced high-density lipoprotein (HDL) cholesterol: fasting HDL cholesterol <1.03 mmol/L in men or 1.29 mmol/L in women, or specific treatment.

Height Assessment
We categorized height into tertiles to illustrate the shape of any associations. However, height diminishes with age due to shrinkage, and height increases over generations.

<table>
<thead>
<tr>
<th>Relative Height Tertile</th>
<th>Short (n = 952)</th>
<th>Middle (n = 953)</th>
<th>Tall (n = 955)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yrs); mean</td>
<td>45.9</td>
<td>45.6</td>
<td>45.9</td>
</tr>
<tr>
<td>Height (cm); mean</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>159</td>
<td>165</td>
<td>171</td>
</tr>
<tr>
<td>Women</td>
<td>147</td>
<td>153</td>
<td>159</td>
</tr>
<tr>
<td>Education; (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary or less</td>
<td>41</td>
<td>31</td>
<td>29</td>
</tr>
<tr>
<td>Secondary</td>
<td>46</td>
<td>51</td>
<td>48</td>
</tr>
<tr>
<td>Matriculation or greater</td>
<td>13</td>
<td>18</td>
<td>23</td>
</tr>
<tr>
<td>Job activity; (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No job</td>
<td>29</td>
<td>30</td>
<td>28</td>
</tr>
<tr>
<td>Sedentary</td>
<td>13</td>
<td>16</td>
<td>18</td>
</tr>
<tr>
<td>Mild</td>
<td>22</td>
<td>23</td>
<td>25</td>
</tr>
<tr>
<td>Moderate</td>
<td>30</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>Heavy</td>
<td>6</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Marital status; (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>80</td>
<td>83</td>
<td>84</td>
</tr>
<tr>
<td>Single/widowed/divorced</td>
<td>20</td>
<td>17</td>
<td>16</td>
</tr>
<tr>
<td>Ever use alcohol; (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>37</td>
<td>36</td>
<td>38</td>
</tr>
<tr>
<td>No</td>
<td>63</td>
<td>64</td>
<td>62</td>
</tr>
<tr>
<td>Smoking status; (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>75</td>
<td>73</td>
<td>74</td>
</tr>
<tr>
<td>Ex-smoker</td>
<td>6</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Current smoker</td>
<td>19</td>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td>Leisure exercise in past month; (%)</td>
<td>45</td>
<td>41</td>
<td>42</td>
</tr>
<tr>
<td>Yes</td>
<td>55</td>
<td>59</td>
<td>58</td>
</tr>
</tbody>
</table>

We therefore created tertiles of relative height for age and sex using internally generated sex-specific curves with linear, square and cubic terms of age.

Statistical Analysis
The association of the metabolic syndrome and its components with height was assessed using logistic regression adjusted for potential confounding factors. We computed odds ratios (ORs) and 95% confidence intervals (CIs). Men and women were analyzed together. If the effect of height differed between the sexes (ie, a model with an interaction term for height tertile with sex had a smaller Akaike Information Criteria), men and women were also analyzed separately.

Potential confounding factors were age, education, job activity, marital status, use of alcohol, smoking and exercise, as categorized as in Table 1. First we adjusted for age and sex (model 1). Next, we added all other confounders (model 2). Finally, we also adjusted the 4 appropriate components of the metabolic syndrome for central obesity (model 3).

RESULTS
Of the 2900 participants examined, 5 were excluded because of current pregnancy, and 35 had incomplete data (4 missing the metabolic syndrome, 3 missing height, and 28 missing other confounders). Taller people had more education and less physically active jobs (Table 1). Older people more often had the metabolic syndrome (Table 2). Across all age groups, people with the metabolic syndrome were generally taller (Fig. 1), with the height difference possibly being greater in men. In the adjusted models, there was no evidence of different associations between height tertiles and the metabolic syndrome by sex. However height consistently had a stronger association with central obesity in men (OR = 2.80; 95% CI = 1.98–3.97 for tallest compared with shortest tertile) than women (1.63; 1.19–2.23).

In all models, tallness was associated with the metabolic syndrome, central obesity, and reduced HDL cholesterol (Table 3). Increased blood pressure and raised fasting plasma glucose had little relation with height until adjustment

| TABLE 1. Sample Characteristics by Age- and Sex-Adjusted Relative Height Tertiles |
|-----------------------------------|----------------|----------------|---------------|
| Age (yrs); mean                   | 45.9           | 45.6           | 45.9          |
| Height (cm); mean                 |                |                 |               |
| Men                               | 159            | 165            | 171           |
| Women                             | 147            | 153            | 159           |
| Education; (%)                    |                |                 |               |
| Primary or less                   | 41             | 31             | 29            |
| Secondary                         | 46             | 51             | 48            |
| Matriculation or greater          | 13             | 18             | 23            |
| Job activity; (%)                 |                |                 |               |
| No job                            | 29             | 30             | 28            |
| Sedentary                         | 13             | 16             | 18            |
| Mild                              | 22             | 23             | 25            |
| Moderate                          | 30             | 26             | 26            |
| Heavy                             | 6              | 4              | 4             |
| Marital status; (%)               |                |                 |               |
| Married                           | 80             | 83             | 84            |
| Single/widowed/divorced           | 20             | 17             | 16            |
| Ever use alcohol; (%)             |                |                 |               |
| Yes                               | 37             | 36             | 38            |
| No                                | 63             | 64             | 62            |
| Smoking status; (%)               |                |                 |               |
| Never                             | 75             | 73             | 74            |
| Ex-smoker                         | 6              | 7              | 6             |
| Current smoker                    | 19             | 20             | 19            |
| Leisure exercise in past month; (%)| 45 | 41 | 42 |
| Yes                               | 55             | 59             | 58            |

| TABLE 2. Prevalence of the Metabolic Syndrome by Age and Sex |
|--------------------------|----------------|----------------|---------------|
| Age Group (yrs)          | Men            | Women          |
| 25–29                    | 6              | 0              |
| 30–34                    | 6              | 3              |
| 35–39                    | 9              | 6              |
| 40–44                    | 6              | 7              |
| 45–49                    | 11             | 17             |
| 50–54                    | 19             | 22             |
| 55–59                    | 20             | 40             |
| 60–64                    | 24             | 36             |
| 65–69                    | 25             | 38             |
| 70–74                    | 20             | 46             |
was made for central obesity (model 3), upon which shortness was associated with raised blood pressure and raised fasting plasma glucose. Central obesity in the taller participants confounded the relationships of greater height with lower risk of raised blood pressure and fasting plasma glucose. Adjustment by body mass index (weight [kg]/height squared [m²]) rather than central obesity produced no such confounding effect (data not shown).

DISCUSSION

In a recently developed Chinese population, taller stature was associated with the metabolic syndrome, in part because taller people (especially taller men) were more likely to be centrally obese. Shorter height was not clearly related to increased blood pressure or raised fasting plasma glucose, until after adjustment for the confounding effect of central obesity. Our results from an understudied ethnic group with a different social patterning of cardiovascular disease have some similarities with those from long-term industrialized countries. 

Several limitations bear mention. First, the sample was better educated than the population; bias would occur if taller people without risk factors or shorter people with risk factors were thereby excluded, which we have no reason to expect. The long life expectancy and low rate of cardiovascular disease in Hong Kong make it unlikely that premature adult deaths caused these findings; after about 8 years of follow-up, only 29 (1%) had died of cardiovascular disease. Second, age ranged from 25 to 74 years; however, there was consistency across the age-range (Fig. 1). Third, central obesity was proxied by waist circumference, which could be proportional to height. The association of greater height with the metabolic syndrome and central obesity could then be due to a larger skeleton. However, there is little evidence that waist circumference is proportional to height, that it increases with secular increases in height, or that it increases more with height in men than women. Moreover, such a relationship to be centrally obese.

### TABLE 3. Association Between Relative Height Tertile and the Metabolic Syndrome and its Components

<table>
<thead>
<tr>
<th>Condition</th>
<th>Prevalence (%)</th>
<th>Model*</th>
<th>Middle OR (95% CI)</th>
<th>Tall OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metabolic syndrome</td>
<td>14.8</td>
<td>1</td>
<td>1.78 (1.34–2.36)</td>
<td>1.85 (1.40–2.44)</td>
</tr>
<tr>
<td>Central obesity</td>
<td>25.5</td>
<td>1</td>
<td>1.63 (1.30–2.04)</td>
<td>1.94 (1.56–2.43)</td>
</tr>
<tr>
<td>Increased blood pressure</td>
<td>28.9</td>
<td>1</td>
<td>1.09 (0.87–1.37)</td>
<td>0.91 (0.72–1.14)</td>
</tr>
<tr>
<td>Increased fasting plasma glucose</td>
<td>18.8</td>
<td>1</td>
<td>0.97 (0.76–1.23)</td>
<td>0.74 (0.58–0.94)</td>
</tr>
<tr>
<td>Increased triglycerides</td>
<td>17.5</td>
<td>1</td>
<td>1.23 (0.96–1.57)</td>
<td>1.27 (1.00–1.63)</td>
</tr>
<tr>
<td>Decreased HDL cholesterol</td>
<td>42.8</td>
<td>1</td>
<td>1.31 (1.09–1.57)</td>
<td>1.46 (1.22–1.76)</td>
</tr>
</tbody>
</table>

*Model 1 adjusts for age (5-yr bands) and sex.
Model 2 adjusts for age (5-yr bands), sex, education, marital status, job activity, smoking status, use of alcohol, and leisure exercise.
Model 3 adjusts for age (5-yr bands), sex, education, marital status, job activity, smoking status, use of alcohol, leisure exercise, and the presence of central obesity.

Reference category is shortest height tertile.
between height and waist circumference would invalidate definitions of the metabolic syndrome using single sex-specific waist criteria. Finally, we had one percent missing data for confounders; however, these confounders had little effect on the estimates (Table 3). Repeat analysis with more complete data, but no confounders, gave similar results.

Height is a composite marker of 2 childhood exposures that have different associations with cardiovascular risk. Prepubertal growth (of the legs) is usually protective, whereas relatively greater pubertal growth (of the trunk) appears detrimental.\(^3\)\(^2\)\(^3\)\(^5\)\(^6\)\(^9\)\(^10\)\(^11\)\(^12\) We have no information on body proportions. The possibility that rapid and recent epidemiologic transition resulted in height that largely represented detrimental pubertal growth cannot be excluded; such an effect would be greater in men than in women, because men grow more at puberty.\(^3\)\(^6\)\(^10\)\(^31\) Accelerated postnatal growth is also associated with obesity\(^3\)\(^2\) and cardiovascular risk,\(^3\)\(^3\)\(^4\) although the mechanisms are not well understood. Rapid intergenerational improvements in childhood environments could promote such accelerated growth\(^3\)\(^5\) relative to the mother’s constitution. These participants are taller than previous generations.\(^3\)\(^6\)\(^36\) Currently, there are larger intergenerational increases in height in Hong Kong than in long-term industrialized countries.\(^3\)\(^7\)

We speculate that height represents both protective and detrimental growth, making the effect of height on cardiovascular risk contextually driven, depending on which factor dominated in a specific context. Thus, the social production of cardiovascular risk may also depend on the population’s socioeconomic history.

ACKNOWLEDGMENTS

The Hong Kong Cardiovascular Risk Factor Prevalence Study Steering Committee consisted of the following members: E. D. Janus (Chairman), C. S. Cockram, R. Fielding, A. J. Hedley, P. Ho, C. P. Lau, M. Lo, S. L. Lo, P. L. Ma, J. R. C. Maserei, Y. T. Tai, B. Tomlinson, S. P. Wong, J. L. F. Woo. The authors would like to thank the late M. R. Janus, survey center nurse coordinator; S. F. Chung for her assistance in recruitment and telephone interview coordination; and S. T. S. Siu for assistance in data processing.

REFERENCES


Readers with a medical background will immediately recognize “Austin Flint” as the name for the rumbling murmur of aortic regurgitation. Flint was a physician who described this murmur in 1866, and his interpretation has stood the test of time. Austin Flint achieved distinction in other respects as well: he helped to establish 2 major medical schools, served on the faculties of 6 medical schools, wrote a major medical text, and was the first American to address the British Medical Society.

So why should epidemiologists take notice of Austin Flint? In October 1841, an epidemic of typhoid fever broke out in the tiny village of North Boston, 8 miles south of Buffalo, New York. Among its 43 inhabitants, 23 were infected, and 10 died. Thirty-one-year-old physician Austin Flint was sent by the county superintendent of the poor to investigate.

Flint found that a traveler from an area with typhoid was overcome by illness and stopped at the local inn. He died 4 weeks later. The first North Boston case occurred 21 days after his arrival. Many local residents, including the innkeeper and his family, had visited the sick traveler. Flint inferred that the illness had been transmitted by those contacts, rather than through miasmas, as was the prevailing theory.\(^1\)

There was another theory among the villagers. The innkeeper was feuding with a man named Stearns, and had denied Stearns access to the inn’s well. As it happened, Stearns and his family were among those unaffected by the outbreak. This led some in the village to believe that Stearns had poisoned the well. Flint called in a medical examiner, who confirmed that typhoid—and not poisoning—was the cause of the deaths. Flint’s 1845 paper attributed the typhoid transmission to contagion, and became important in opposing the prevalent miasmatic theory of contagion.

Twenty-eight years later, Flint returned to the possibility that the well had in fact been the source of the outbreak—perhaps after reading William Budd’s observations of typhoid fever transmission through drinking water. Flint concluded that, indeed, the inn’s well must have been contaminated by the sick traveler’s excrement. Flint wrote, “It can hardly be doubted that the exemption of the family of Stearns was due to the animosity of the innkeeper, which led the latter to prohibit the use of his well, and compelled Stearns to dig a well of his own.” (italics in the original). Flint humbly reported his revised conclusions at the very first meeting of the American Public Health Association in 1872.\(^2\)

REFERENCES

To the Editor:

There is a strong relationship between a woman’s weight just prior to pregnancy, or in early pregnancy, and her risk of gestational diabetes mellitus, pre-eclampsia, placental dysfunction, and adverse perinatal events. In the past decade, the prevalence of obesity among pregnant women may have increased. Less is known about ethnic-specific temporal trends in maternal weight and the risk of obesity in pregnancy. Examining these trends is the goal of the Trends in Obesity in Pregnancy Study, which is a population-based study, derived using an administrative database, as described elsewhere.

Briefly, since late 1993, standardized maternal serum screening has been available to all women in Ontario, at 15 to 20 weeks’ gestation, through their physician or midwife. Steady uptake has occurred from 1994 onward. Data about maternal date of birth, parity, ethnicity, and weight are recorded in a standardized fashion on a form, completed by the caregiver at the time of the screening.

We defined the mutually exclusive overweight, obese, and severely obese states as presented in the footnote to the Table. The prevalence of each state in the years 1994 and 2000 were compared using polytomous logistic regression analysis. The resulting odds ratios were adjusted for maternal age, ethnicity, parity, gestational age at weight measurement and presence of a multifetal pregnancy.

A total of 369,740 women underwent maternal serum screening, at an average of 16.3 weeks’ gestation (Table 1). In each ethnic group, the median gravidity was 2 and the median parity 1. Among all women, as well as within each of the 4 ethnic groups, there was a significant increase in mean maternal weight over time, based on a time-series model incorporating a linear trend term (P < 0.001; Fig. 1 available with the electronic version of this article at www.epidem.com).

Among all women, the prevalence of the overweight state increased from 9.6% in 1994 to 11.4% in 2000 (adjusted odds ratio = 1.37; 95% confidence interval = 1.31–1.43). The prevalence of obesity rose from 4.0% to 5.6% (1.7; 1.6–1.8), and severe obesity increased from 0.99% to 1.5% (1.8;1.6–2.0).

As a limitation, height was not recorded in TOPS, and body mass index was not calculated. At the same time, maternal weight approximates BMI quite well, and by limiting our assessment of mean weight to each ethnic group, we were able to control, to a degree, for the variation in height between racial groups. Since maternal weight was recorded at around the same gestational age, and we further adjusted for gestational age, ethnicity and other factors, variance and confounding were likely minimized.

The data from this study are consistent with other studies of temporal weight trends in pregnancy. These data also parallel rising trends of obesity among children and adolescents in the United Kingdom and United States. Together, they emphasize that the rate of obesity and morbidity in pregnancy is likely to increase over the next decade.

This study provides novel information about weight trends according to broad ethnic groups, and the results highlight the need to consider preventive strategies among all ethnic groups, including those traditionally deemed to be at “low risk,” such as women of Asian ancestry.

Given that weight loss during pregnancy may have untoward effects, preventing maternal obesity requires action before a woman reaches her reproductive years. Promoting controlled calorie-energy intake and higher energy expenditure in childhood or adolescence may be the most sensible approach. Ongoing epidemiologic surveillance is needed to detect any change in the incidence of pre-eclampsia, gestational diabetes mellitus, operative or preterm delivery or perinatal mortality, also stratified by maternal

### Table 1. General Characteristics of Participants at the Time of Maternal Serum Screening in Ontario, 1994–2000

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>All Women</th>
<th>White</th>
<th>Black</th>
<th>First Nations</th>
<th>Asian</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years); mean ± SD</td>
<td>29.0 ± 4.0</td>
<td>29.3 ± 4.0</td>
<td>29.1 ± 4.0</td>
<td>29.4 ± 4.0</td>
<td>29.2 ± 4.0</td>
<td>29.0 ± 4.0</td>
</tr>
<tr>
<td>Gestational age at which weight was measured (weeks); mean ± SD</td>
<td>16.2 ± 1.1</td>
<td>16.2 ± 1.1</td>
<td>16.2 ± 1.1</td>
<td>16.2 ± 1.1</td>
<td>16.2 ± 1.1</td>
<td>16.2 ± 1.1</td>
</tr>
<tr>
<td>Gestational weight (kg); mean ± SD</td>
<td>69.0 ± 14.4</td>
<td>69.0 ± 14.4</td>
<td>69.0 ± 14.4</td>
<td>69.0 ± 14.4</td>
<td>69.0 ± 14.4</td>
<td>69.0 ± 14.4</td>
</tr>
<tr>
<td>Obesity†</td>
<td>12.3</td>
<td>5.7</td>
<td>15.7</td>
<td>13.5</td>
<td>3.0</td>
<td>4.6</td>
</tr>
<tr>
<td>Overweight*</td>
<td>30.7</td>
<td>25.3</td>
<td>32.8</td>
<td>30.7</td>
<td>31.2</td>
<td>31.2</td>
</tr>
<tr>
<td>Severe obesity‡</td>
<td>5.7</td>
<td>3.3</td>
<td>7.0</td>
<td>5.2</td>
<td>0.62</td>
<td>0.56</td>
</tr>
</tbody>
</table>

†Defined as a maternal weight >85th to 95th centile (79.0–92.2 kg) among all women in 1994.
‡Defined as a maternal weight >95th centile (92.3–112.2 kg) among all women in 1994.
*Defined as a maternal weight >99th centile (112.3 kg) among all women in 1994.
weight or BMI before, or in early pregnancy.

To the Editor:

In the very nice and well-done study, Baylin et al found that transient exposure to coffee may trigger myocardial infarction (MI), especially in patients with occasional intake of coffee (≤1 cup/d). Interestingly, they found no risk elevation in heavy coffee drinkers. This is remarkable, as most studies until now examined the effect of coffee on cardiovascular function in patients without considering the amount of coffee consumed. In our opinion, this aspect should be taken into consideration and highlighted. There is a lot of speculation about how coffee may trigger MI but not why heavy coffee drinkers are protected.

We recently showed that coffee and caffeine similarly increase sympathetic activity as measured directly by microneurography. We found that while coffee increased blood pressure in nonhabitual drinkers, the effect was blunted in habitual drinkers, despite the same sympathetic activation. Caffeine infusion similarly increased sympathetic nervous activity and blood pressure in habitual and nonhabitual coffee drinkers. The lack of blood pressure response after coffee in habitual drinkers may explain why, in the study of Baylin et al, heavy drinkers are less likely to develop MI. Moreover, we were able to demonstrate that coffee blunts cardiovascular response to mental stress in habitual coffee drinkers, whereas nonhabitual drinkers showed a stress-induced response of systolic blood pressure.

The case–control study by Baylin et al adds important information to the hypothesis that habitual coffee drinkers are somehow protected against the cardiovascular effects of coffee (and its main component, caffeine). These findings support not only our data, but also a previous epidemiologic study not mentioned in the article. When interpreting data on coffee, the amount of coffee drinking should always be taken into account.

Baylin et al hypothesized that the increase in MI risk is mainly attributed to the serum–caffeine concentration. Since decaffeinated coffee also increases blood pressure in nonhabitual coffee drinkers, it has to be stressed that substances other than caffeine may be responsible for blood pressure elevation and increased risk of MI in nonhabitual coffee drinkers.

Our above-mentioned studies showed that both decaffeinated coffee and caffeine resulted in a similar time-by-condition interaction for total sympathetic nerve activity in nonhabitual coffee drinkers. Thus, substances other than caffeine are likely to be responsible for the stimulating effect seen in this subset of individuals.

REFERENCES


REFERENCES

Coffee and Myocardial Infarction

To the Editor:

The role of daily activities and events in triggering myocardial infarction (MI) remains inconclusive. Baylin et al1 conducted a study supporting the idea that coffee intake may trigger MI. However, they recognized the possibility of temporal confounding by other external triggers such as smoking, physical, or sexual activity. Additional confounders may be emotional or meteorological stresses,2 but the greatest confounding could be through intake of various substances other than coffee, possibly taken at the same time as the coffee. Food ingestion precedes approximately 8% of MIs3 and has been verified as a trigger of acute coronary syndrome by case–crossover methodology.4 Cold drinks5 and alcoholic drinks6 have also been suggested as triggers of sudden cardiac events. Finally, urination and defecation can follow coffee consumption and have been implicated in the triggering of MI.7

Similar to the present study,8 Selb Šemerl and Šelb estimated the hazard period for a cardiac event as 1 hour after ingesting coffee.5 In their study, regular physical activity was not protective, but rather increased the likelihood of sudden cardiac death. It may be that coffee intake predisposes to fatal cardiac arrhythmias in people who are susceptible but apparently healthy (engaging in regular physical activity), whereas in those with a higher risk profile (who have presumably already developed atherosclerosis), coffee may instead trigger typical coronary atherothrombosis.

This is a continuing controversy in the present study.1 The risk of MI was greater among those with 3 or more risk factors for coronary disease; however, when the risk factors were examined separately, the risk of MI was lower among those with hypertension, hypercholesterolemia, or history of previous angioplasty compared with those not having these risk factors.3 Recently, it has been reported that coffee intake increases the risk of nonfatal MI only among individuals with “slow” caffeine metabolism, such as carriers of the variant cytochrome P450 1A2*1A, but not among “rapid” caffeine metabolizers who are homozygous for the cytochrome P450 1A2*1A allele.8 An acute cardiac event is often multifactorial, and the potential of coffee to trigger a specific event should be further defined in relation to an individual’s demographic and risk factor characteristics. A possibility of diverse susceptibility to a particular external trigger according to such characteristics has previously been suggested.2,9

The study by Baylin and colleagues10 that people with light or occasional coffee intake may be at higher risk of MI than those with higher intake. Baylin et al reasonably proposed that the triggering effect of coffee could be greater in the morning due to lower plasma caffeine even in heavy drinkers. Moreover, given the fact that coffee is more often used in the morning, these 2 correlated effects may contribute to the morning peak in incidence of MI. Coffee may also contribute to the slight Monday excess in MIs and other acute cardiac events.11 It had been proposed that this excess is an artifact of events with uncertain dates being coded as taking place on Mondays,12 but this has not been supported by more recent meta-analyses.11 In addition to other putative explanations,13 the start of the work week may be accompanied by a relative excess in coffee intake after 2 days of weekend decrease in consumption and tolerance, which could cause a more marked coffee-related sympathetic surge. This sounds more plausible than a Monday excess related to alcohol ingestion,11 since peak consumption of alcohol takes place on Friday and Saturday.13 However, this needs further exploration.

Viktor Čulić
University Hospital Split
Split, Croatia

REFERENCES


The authors respond:

Hammer et al1 provides important physiological explanations for our epidemiological findings2 as well as relevant information that should be taken into account in future studies. They also mention the potential effect of decaffeinated coffee, which would support the role of substances other than caffeine in the cardiovascular effects attributed to coffee drinking. Unfortunately, the small number of decaf coffee drinkers in our Costa Rican population prevented us from evaluating its effects.

Čulić3 suggests a series of other potential triggers that could have confounded the temporal association found between coffee intake and myocardial infarction in our study. In our analysis, we carefully evaluated potential confounders such as smoking, heavy physical exertion, sexual intercourse, and anger. Other plausible confounders will remain as alternative explanations to the findings since we do not have the data to either prove or refute their effect. Further studies are needed to understand the important effects of these other triggers.

© 2007 Lippincott Williams & Wilkins
We agree with Čulic in that other issues related to chronobiological patterns need to be studied in more detail in the context of trigger research.

Ana Baylin
Department of Community Health
Brown University
Providence, RI
ana_baylin@brown.edu

Sonia Hernandez-Diaz
Department of Epidemiology
Harvard School of Public Health
Boston, MA

Edmond K. Kabagambe
Department of Epidemiology,
School of Public Health
University of Alabama at Birmingham
Birmingham, AL

Xinia Siles
Centro Centroamericano de Población
Universidad de Costa Rica
San Pedro, Costa Rica

Hannia Campos
Department of Nutrition
Harvard School of Public Health
Boston, MA
Centro Centroamericano de Población
Universidad de Costa Rica
San Pedro, Costa Rica

REFERENCES

Calculating Disease Burden

To the Editor:

With interest, I read the overview of methods for calculating the burden of disease due to specific risk factors given by Steenland and Armstrong.1 Although I agree that epidemiologists should try to make the consequences of their research work more obvious to potential users, such as standard setting committees, or industrial injury advisory councils, I’d like to point out that an uncritical use of the methods suggested may lead to a biased discussion and evaluation of the effects of exposures. The authors noted on p. 513 that the attributable fraction among the exposed is included in the “etiologic fraction,” which is the percentage of subjects causally affected by exposure (given the implicit assumption that exposure is never preventive). It follows that the attributable fraction may serve only as an underestimate of the etiologic fraction. An uncritical use of the attributable fraction therefore may produce a bias against an exposed worker if the attributable fraction is used to decide against compensation.2

The situation is worse with years of life lost and disability-adjusted years of life lost for specific causes of death. The authors noted on p. 517, with a view to the fundamental paper by Robins and Greenland,3 that they rely on the assumption that among the exposed who succumb to the disease in question, vulnerability to other causes of death is not also heightened. This is an assumption usually not met in cohort mortality studies. And even under the supposition of exposure being never preventive, the estimators suggested by the authors may grossly overestimate the true years of life lost due to exposure in realistic scenarios.4

How should politicians use these estimates if it can be shown easily that the bias away from the null may be greater than 70% in concrete applications?5,6 Therefore, the suggested estimators should not be used as “breadbrush descriptors of disease burden” (p. 517) without quantifying the potential bias, at least its direction.

I’d like to add that the authors missed the concept of average attributable fractions6,7 that provides a solution to the problem of dividing the total proportion of cases attributable to the entire set of exposures under study into exposure-specific components. Moreover, the authors did not mention important alternative concepts to years of life lost, such as risk or rate advancement periods6,8 average age at first occurrence10 and g-estimation.6

Peter Morfeld
Institute for Occupational Medicine
University of Cologne
Cologne, Germany
peter.morfeld@rag.de

REFERENCES

The authors respond:

We thank the author of this letter1 for his comments. We agree that epidemiologists should be aware of the assumptions on which their methods rest and the biases to which they are subject. However, we do not agree that epidemiologists should decline to use estimators because they are subject to unquantifiable biases. Indeed, such a stricture would leave little that epidemiologists could estimate, since unquantifi-
fiable residual confounding potentially affects virtually all epidemiologic estimates. Part of the task of epidemiology is to assess the plausible extent of such biases.

We also agree that the assumptions necessary for the estimators of years-of-life-lost are unlikely to be met exactly. Nevertheless, the key assumption, in addition to those required for validity of attributable fraction, is transparent: we assume the person who got the disease because of the exposure would (if unexposed) have lived free of the disease for at least the duration observed for a similar person in the reference population. Plausible extent of departures from this assumption can be assessed informally and noted in reporting. Finally, we thank the author of the letter for bringing to our attention references to other useful publications on this topic.

Ben Armstrong
Public and Environmental Health Research Unit
London School of Hygiene and Tropical Medicine
London, United Kingdom

Kyle Steenland
Rollins School of Public Health
Emory University
Atlanta, GA
nsteenl@spherit.emory.edu

REFERENCE

Breast Cancer Trends

To the Editor:

Tarone has reported a decline in the incidence of invasive breast cancer since 1990 in American women ages 40–49 years, coincident with a rapid increase in the incidence of in situ breast cancer. Tarone suggests that one or more heretofore unrecognized environmental factors (perhaps related to industrialization) that have changed in recent years could be responsible for the decreasing incidence of invasive breast cancer because (1) the prevalence of risk factors known to be associated with the incidence of invasive breast cancer has not been increasing and (2) the rising use of mammography in American women in their 40s (reflected in the rising incidence of in situ breast cancer) possibly could lead to an increase, not a decrease, in the reported incidence of invasive breast cancer, given that mammography may identify some invasive lesions that would not have become evident during a given woman’s lifetime.

I suggest that the basis for the decreasing incidence of invasive breast cancer in 40- to 49-year-old American women since 1990 may lie instead in the recognition and treatment of an increasing number of women in this age range with in situ breast cancer. The goal of treating in situ breast cancer, after all, is the prevention of invasive breast cancer, and it is indeed likely that a substantial proportion of women with in situ breast cancer would develop invasive disease within a few years time if not treated. Similarly, the contemporaneous increases in the incidence of diagnosed in situ cervical cancer and decreases in the incidence of invasive cervical cancer serve as an important piece of evidence in support of the efficacy of cervical screening. These trends have not been viewed as a stimulus to identify environmental factors that protect against the development of invasive cervical cancer, but rather as evidence of a beneficial effect of the recognition and treatment of precursors of this condition. I believe that the opposing time trends in the incidence of in situ and invasive breast cancer in American women in their forties can be interpreted in the same way.

Noel S. Weiss
Department of Epidemiology
School of Public Health
University of Washington
Seattle, WA
nweiss@u.washington.edu

REFERENCES

The author replies:

Weiss suggests that the declining incidence of invasive breast cancer, which we reported recently, could be due to increased recognition and treatment for in situ cancer. Closer examination of trend data indicates otherwise. Improvements in medical practice usually are manifested in age-period-cohort analyses as changes in the calendar period pattern of disease risk, because such improvements tend to affect the health of people simultaneously in multiple age groups when they are introduced. For example, greater use of mammography led to remarkably similar calendar period increases of in situ breast cancer incidence rates after 1982 for women in all decades of age between 20 and 69 years-of-age. For both invasive breast cancer incidence rates and breast cancer mortality rates, however, a change in the birth cohort pattern of risk is apparent for white and black U.S. women, namely, moderation in the birth cohort risk for women born after 1945.

Increased recognition and treatment of in situ breast cancer cannot explain the decreased breast cancer risk apparent in the Figure for women born after 1945 because the decreases in the youngest age groups began before the widespread use of mammography. In fact, the decreases in breast cancer rates in the 20- to 24-, 25- to 29-, and 30- to 34-year age groups all began before the rapid expansion in the use of mammography in the mid-1980s. To account for the trends in the Figure, a beneficial change in medical practice would have to have been introduced first to women 20–24 years-of-age in the late 1960s, then to women 25–29 years-of-age 5 years later, then to women 30–34 years-of-age after another 5 years, and so on. Changes in the birth cohort pattern of risk like that shown in the Figure provide evidence that there were beneficial changes in underlying risk or protective factors for breast cancer beginning at a relatively early age for U.S. women born after 1945.

In age-period-cohort analyses of U.S. cervical cancer mortality rates, the impact of the introduction of the Pap test was manifested as a decrease in the calendar period slope beginning in the
early 1960s. Estimated birth cohort effects from these analyses indicated that the birth cohort risk of cervical cancer increased, rather than decreased, for women born after 1945, likely reflecting changing sexual mores in the 1960s and 1970s. It is sobering to contemplate the cervical cancer epidemic that might have been, were it not for the introduction of the Pap test in the 1950s.

Robert E. Tarone
International Epidemiology Institute
Rockville, MD
bob@iei.ws

REFERENCES

FIGURE. Age-specific breast cancer mortality rates from 1950 through 2002 for U.S. white women by year of birth from SEER data. The dotted vertical line indicates the 1946 birth year.