Non-differential misclassification and bias towards the null: a clarification

Editor—In a recent paper, Sorahan and Gilthorpe use simulation studies to produce estimates of risk ratios (RRs) with data that are misclassified randomly and independently of disease state.[1] They show that these estimates can be more extreme than either RR_T, the true risk ratio (with no random variation and no misclassification), or RR_{NM}, the risk ratio with random variation but no misclassification, called "actual risk ratio" by Sorahan and Gilthorpe.[1] This is an important point for readers to appreciate. Their report prompts several observations on the general topic of non-differential misclassification in either cohort (their example) or case-control studies.

(1) The most important and the simplest point is that non-differential misclassification of a binary exposure (exposed or not) and a perfectly classified binary outcome (diseased or not) does indeed produce a *bias* toward the null. Always. (In one special case, the effect of misclassification is bias *beyond* the null. This reversal of the direction of association can occur only when the measurement is so bad that the sum of specificity and sensitivity is below 1.) *Bias* refers to a *systematic tendency* and not to a particular result. Here, the bias is the difference between the expected value (average over infinitely many hypothetical replications) of an estimator of the risk ratio calculated with misclassified exposure and the RR_T, the expected value of the risk ratio estimator when there is no error.

By calculation of the value of the risk ratio with the specified rates of misclassification and of disease in the exposed and unexposed populations, one can establish the existence and magnitude of bias. For set 1 of Sorahan and Gilthorpe's simulation, the classification had 90% sensitivity and specificity of exposure and probabilities of disease of 0·0075 and 0·0050 for exposed and unexposed people, respectively.[1] Define RR_{EM} as the expected value of the risk ratio when exposure is misclassified and $\widehat{RR}_{EM}$ as an estimator of RR_{EM} called "apparent risk ratio" by Sorahan and Gilthorpe.[1] Under these assumptions, as sample sizes increase, $\widehat{RR}_{EM}$ converges to the value RR_{EM}.[2]

$$\frac{\dfrac{(0{\cdot}9 \times 0{\cdot}0075) + (0{\cdot}1 \times 0{\cdot}005)}{(0{\cdot}9 \times 5000) + (0{\cdot}1 \times 5000)}}{\dfrac{(0{\cdot}9 \times 0{\cdot}005) + (0{\cdot}1 \times 0{\cdot}0075)}{(0{\cdot}9 \times 5000) + (0{\cdot}1 \times 5000)}} = 1{\cdot}38.$$

The value RR_{EM} is also very near the expected value of $\widehat{RR}_{EM}$ for large samples, and, therefore, one should expect that the median of the distribution of $\widehat{RR}_{EM}$ from the simulation should be near $RR_{EM} = 1{\cdot}38$. Indeed, 1·38 is exactly the median value reported in Sorahan and Gilthorpe's table 2.[1] Repeating this exercise for the other situations considered by Sorahan and Gilthorpe yields RR_{EM}s that are also below the RR_T of 1·5 and very close to the medians of $\widehat{RR}_{EM}$ reported in the table.[1] The fact that $1 < RR_{EM} < RR_T$ in each case proves that the bias in these situations is towards the null.

Sorahan and Gilthorpe note that, in previous work, "both disease outcome and exposure misclassification were assumed to operate on a proportionate rather than a random basis".[1] The reason for this "assumption" is clear: these simple calculations can show, *without simulations*, the magnitude of the bias from random misclassification.

(2) The study of Sorahan and Gilthorpe shows well how the systematic and the random components, which are quite distinct in principle, may interact in practice. The fact that $\widehat{RR}_{EM}$ was above 1·5 in some simulations shows the impact of random variation counteracting a systematic tendency. The combination of the two components also raises an interesting point about the theoretical treatment of misclassification in the epidemiological literature. Sometimes non-differential misclassification is treated as a process—that is, misclassification is not more likely on average in cases or controls—and sometimes as the realisation in the data—that is, the same fraction of cases and controls were misclassified in the study at hand. When the misclassification is treated as a process, bias, estimated by comparison of $\widehat{RR}_{EM}$ with RR_T as in column 9 of table 1 of Sorahan and Gilthorpe,[1] is the concern. By contrast, when differential misclassification refers to the data, comparison of $\widehat{RR}_{EM}$ with $\widehat{RR}_{NM}$ (column 10) is the issue. When the investigator cannot calculate empirical misclassification percentages, one must judge whether the process is non-differential.

The distinction between a misclassification process and the empirical misclassification in a study provides another way to understand the simulation results that Sorahan and Gilthorpe find disturbing. How do we explain the fact that for many of the realisations there is a stronger effect in the misclassified data than in the correctly classified data ($\widehat{RR}_{EM} > \widehat{RR}_{NM}$)? In these instances, the misclassification actually was differential *in the data*. That is, even when the classification *process* yields errors for cases and non-cases equally often in the long run, the *empirical* misclassification in any given study can easily be differential simply due to chance.

A hypothetical example may help. Out of 5000 exposed and 5000 unexposed subjects, the expected numbers of cases are 37·5 and 25, respectively, implying an RR_T of 1·50. By chance, 40 and 22 could be observed in a particular study and would yield a value of 1·82 for the $\widehat{RR}_{NM}$. How would the effects of non-differential and differential misclassification *in the data* affect these results? If the unexposed subjects from the study were classified correctly and exactly 10% of exposed cases and non-cases were misclassified as unexposed, the apparent numbers of exposed and unexposed cases would be 36 and 26, and the risk ratio estimate would be 1·69, less than the 1·82 already calculated from the data correctly classified but subject to random variation. Thus, misclassification that is non-differential in the data results in a reduction in the rate ratio that would be obtained in the study. Still, the estimate from the misclassified data is greater than the RR_T value of 1·5, as a consequence of random variability. Actual misclassification among exposed subjects of 15% for non-cases and 10% for cases are empirically differential but realistic for an underlying misclassification process that is random and non-differential; these would yield an $\widehat{RR}_{EM}$ 1·87, greater than the $\widehat{RR}_{NM}$ of 1·82 and greater than the RR_T of 1·5, even though the underlying process is non-differential.

Our example shows that random variation in a non-differential classification process can lead to an effect that is greater than either the true RR_T value or the value that would result from the study with perfect classification. This occurs because, by chance, the non-differential process may result in differential misclassification in the data. Of course, statistical theory and the simulation results show that exaggeration of the risk ratio is less likely than attenuation.

(3) Although the simulations of Sorahan and Gilthorpe[1] do not disprove the well established bias towards the null in the classic binary situation they studied, there are special circumstances in which the bias towards the null ought not be invoked without further examination of the statistical model and the likely error structure. Several papers published since 1990 have shown that there are special circumstances where there is a bias towards exaggeration of effects. Dosemeci *et al* identified a scenario where non-differential misclassification of exposure more often than not leads to an overestimate of the odds ratio in an *intermediate* exposure category when there are more than two exposure levels. Other papers that have appeared since the textbooks cited by Sorahan and Gilthorpe[1] were published during the 1980s, have identified circumstances where an overestimate is more likely than an underestimate. These include particular forms of non-differential misclassification when an exposure is not binary,[3,4] when grouping has occurred,[5,6] or when the errors in a continuous exposure are correlated with their true value.[7]

Ultimately, investigators must interpret a specific estimate, such as 1·37, without the information to help us distinguish between the effects of sampling variation and of misclassification. If we then posit neither misclassification nor other biases, we can infer that the confidence interval covers the true value of the parameter with the specified probability. That is, sometimes RR_T will be higher than 1·37 and sometimes lower, and sometimes the confidence interval will not cover the parameter value. The procedure we used to obtain our estimate and confidence interval will perform as expected. On the other hand, if we posit an exposure classification process that has a probability of error that is equal for cases and for non-cases, then we ought to infer that the estimate of 1·37 is *more likely* to fall below the true value than to exceed it. Further, the confidence interval is shifted too low, may have incorrect width, and will cover the true parameter less often than the specified probability.

Many epidemiological studies contend with non-differential measurement error. Sorahan and Gilthorpe show why the practice of inferring that the true value *must* be above the value estimated with non-differential error in exposure assessment is unwarranted.[1] They correctly state that the estimate *may* exceed the true value even when the misclassification process for a binary exposure is non-differential. Still, the estimate is more likely to fall below the true value.

SHOLOM WACHOLDER
PATRICIA HARTGE
JAY H LUBIN
MUSTAFA DOSEMECI
Epidemiology and Biostatistics Program, National Cancer Institute, Bethesda, MD 20892, USA

Correspondence to: Dr Sholom Wacholder, Epidemiology and Biostatistics Program, National Cancer Institute, EPN 403, 6130 Executive Blvd, Bethesda, MD 20892-7368, USA.

1 Sorahan T, Gilthorpe MS. Non-differential misclassification of exposure always leads to an underestimate of risk: an incorrect conclusion. *Occup Environ Med* 1994;51: 839–40.
2 Loeve M. *Probability theory*. Princeton: D Van Nostrand, 1962.
3 Dosemeci M, Wacholder S, Lubin JH. Does non-differential misclassification always bias a true effect toward the null value? *Am J Epidemiol* 1990;132:746–8.
4 Weinberg CR, Umbach DM, Greenland S. When will non-differential misclassification of an exposure preserve the direction of a trend? *Am J Epidemiol* 1994;140:565–71.
5 Wacholder S, Dosemeci M, Lubin JH. Blind assignment of exposure does not always prevent non-differential misclassification. *Am J Epidemiol* 1991;134:433–7.
6 Flegal KM, Keyl PM, Nieto FJ. Differential misclassification arising from non-differential errors in exposure measurement. *Am J Epidemiol* 1991;134:1233–44.
7 Wacholder S. When measurement errors correlate with truth: surprising effects of non-differential misclassification. *Epidemiology* 1995;6:157–61.

Author's reply—Our short report on the properties of non-differential misclassification of exposure, as judged by computer simulations, has prompted Wacholder *et al* to make several useful observations.[1] These observations include a restatement of what we judged to be the "more important" feature of the simulations. We concluded (to paraphrase) that for any particular epidemiological study that investigates a causal risk factor and in which each study subject had the *same probability* of being misclassified (with respect to a single binary exposure variable), it would be *incorrect* to infer that the measure of effect obtained from the study—for example, relative risk or rate ratio—could *only* be increased if more reliable information were to be obtained such that all misclassification could be removed. We are pleased to learn that Wacholder *et al* are of the opinion "that this is an important point for readers to appreciate". We did not find those results of the computer simulations that supported this conclusion to be "disturbing"; they seemed to us to be intuitively obvious. What disturbed us was the fact that many researchers are convinced that the removal of non-differential misclassification of exposure from their studies can only increase the point estimate of relative risk (or rate ratio).

Why is our conclusion so little known? We have three possible explanations; all could be prompted by the comments of Wacholder *et al*. It may be because of confusion about the definition of non-differential misclassification. We chose the definition that "all exposed and non-exposed subjects have the same *probability* of being misclassified (these two probabilities may be different, one must be not zero)". Wacholder *et al* describe this as misclassification "treated as a process". They note that non-differential misclassification may also be defined in terms of "realisation" in a given data set—that is, the same fraction of diseased and non-diseased subjects were, in fact, misclassified. The first definition seems more relevant to study settings. Under the second definition, non-differential misclassification would rarely occur and a researcher would not be aware when it had occurred. (It would never occur when there was an even number of diseased subjects and an odd number of non-diseased subjects!)

A second explanation is the influence of textbook examples in which misclassification is invariably shown to operate on a proportionate rather than a random basis. We choose not to believe that errors are made every *n*th record and prefer to believe that random misclassification is more relevant to study settings.

A third possible explanation is the way in which the word *bias* is interpreted. Sometimes the word is used to indicate a tendency toward a given distortion, and sometimes (perhaps incorrectly) to indicate a distortion that will occur on each and every occasion—for example, in the game of bowls, the oblique course of a bowl due to its lopsided form is said to be due to bias. If the first definition were in universal use, our conclusion would be well known.

Our short report may be viewed as a call for more appropriate interpretation of study findings.[1] The observations of Wacholder *et al* may be viewed in the same light.

TOM SORAHAN
Institute of Occupational Health
MARK S GILTHORPE
Department of Public Health and Epidemiology, University of Birmingham, Edgbaston, Birmingham B15 2TT

1 Sorahan T, Gilthorpe MS. Non-differential misclassification of exposure always leads to an underestimate of risk: an incorrect conclusion. *Occup Environ Med* 1994;51: 839–40.

NOTICES

For a good working life. ICOH'96, the 25th International Congress of Occupational Health. Stockholm, Sweden. 15–20 September 1996.

The Congress will present the latest research discoveries in occupational health as well as provide a forum for exchange of ideas between practitioners and researchers.

This ICOH Congress will be noted by the introduction of new subjects of great concern to the society of today and tomorrow, such as work organisation, psychosocial factors, and gender research. A large number of minisymposia will form a bridge between the more traditional occupational health research and the new challenges of promoting a good working life.

MINI SYMPOSIA

A large part of the conference will be mini symposia arranged by the permanent Scientific Committees, such as "chemicals and allergies", "occupational health nursing in the future", "occupational health in small industries". More than thirty mini symposia will give the participants a unique opportunity to keep up with the rapid developments in occupational health.

MORE THAN 1000 PRESENTATIONS

The Congress also invites participants to present their own papers on many areas of relevance to the working environment. There will be more than 1000 oral presentations and posters.

PROMINENT KEYNOTE SPEAKERS

Eight internationally known scientists are invited as keynote speakers:

- Gender and work. Joan Acker, University of Oregon, USA
- Occupational health—a global perspective. Jerry Jeyaratnam, National University Hospital, Singapore
- Participatory approaches in occupational health. René Loewenson, Zimbabwe
- What can health professionals do to prevent muskuloskeletal disorders? Philippe Mairiaux, Université Catholique de Louvain, Brussels, Belgium
- Working conditions and cardiovascular diseases. Johannes Siegrist, Institut für Medizinisches Soziologie, Düsseldorf, Germany
- Dose concepts in occupational exposure assessments. Thomas J Smith, Harvard School of Public Health, Boston, USA
- Promoting safe behaviour. Carin Sundström-Frisk, National Institute of Occupational Health, Stockholm, Sweden
- Electromagnetic fields and cancer. Gilles Thériault, McGill University, Montreal, Canada.

The conference is sponsored by multinational and Swedish companies. Major sponsors are AMF Trygghetsförsäkring, SmithKline Beecham, Pasteur Mérieux MSD and the Swedish Power Association, Svenska Kraftverksföreningen. Other sponsors are Samhall and SJ, the Swedish State Railways. The official airline is SAS. There will be an exhibition in conjunction with the conference. Companies and organisations interested in taking part should contact the ICOH'96 secretariat.

Authorised press is welcome to cover the Congress.

For further information contact: Arne Wennberg, secretary general ICOH'96, Lars Grönkvist, press officer ICOH'96, Elisabeth Lagerlöf, information ICOH'96, Maud Werner, secretariat ICOH'96, National Institute of Occupational Health S-171 84 SOLNA, Sweden. Tel (+46) 8 730 91 00; Fax (+46) 8 82 05 56.

Broadening the Limits of Occupational Hygiene. 14th Annual Conference of the Australian Institute of Occupational Hygienists. 11-13 December 1995. Adelaide, South Australia.

A three day conference (11–13 December) and two days of continuing education sessions (9–10 December). This conference will run back to back with the 31st Annual Conference of the Ergonomics Society of Australia at the same venue.

An exceptional line up of international and local speakers include

- Professor Morton Lippmann, New York University. Inhalation Toxicology and Regulatory Policy
- Associate Professor Harriet Burge, Harvard University. Bioaerosols
- Professor Jens Rasmussan, Copenhagen University. Human Computer Interaction and Human Error
- Dr Jim Stewart, ex vice president Dupont Canada. Safety Performance and Organisational Change