Nondifferential disease misclassification may bias incidence risk ratios away from the null

Juha Pekkanen\textsuperscript{a,b,*}, Jordi Sunyer\textsuperscript{c}, Susan Chinn\textsuperscript{d}

\textsuperscript{a}Unit of Environmental Epidemiology, National Public Health Institute, P.O. Box 95, 70701 Kuopio, Finland
\textsuperscript{b}Department of Public Health and General Practice, University of Kuopio, Kuopio, Finland
\textsuperscript{c}Unitat de Recerca Respiratòria i Ambiental, Institut Municipal Investigació Médica (IMIM), Barcelona, Catalonia, Spain
\textsuperscript{d}Department of Public Health Sciences, King’s College London, London, United Kingdom

Accepted 13 July 2005

Abstract

Background and Objective: When estimating incidence risk ratios in follow-up studies, subjects testing positive for the disease at baseline are excluded. Although the effect of disease misclassification on estimated incidence risk ratios has otherwise been extensively explored, the effect of disease misclassification at baseline has not previously been analyzed.

Study Design and Setting: The design was theoretical calculations assuming dichotomous disease and a follow-up study with a baseline and a follow-up examination, analyzed using cumulative incidence. Calculations consider nondifferential misclassification of disease mainly at baseline, but no misclassification of exposure.

Results: Nondifferential misclassification of disease at baseline can lead to bias either away or toward null in estimated cumulative incidence risk ratios. This bias is mainly a function of sensitivity at baseline, because imperfect sensitivity leads to failure to exclude all diseased subjects from the follow-up. Imperfect specificity at baseline has less effect. Bias is increased with high true prevalence of disease and low true incidence. Bias is also increased with large differences in true risk ratios at baseline and at follow-up, because observed incidence risk ratios in the presence of misclassification reflect both the true association at baseline and at follow-up.

Conclusion: Nondifferential disease misclassification at baseline examination of a follow-up study can lead to over- or underestimation of the cumulative incidence risk ratios. The bias can be substantial for disease with low incidence and high prevalence, such as asthma or myocardial infarction. The results underscore the need to select a highly sensitive test for disease at baseline to exclude all diseased subjects from the follow-up. © 2006 Elsevier Inc. All rights reserved.

Keywords: Incidence; Follow-up study; Bias; Disease misclassification; Measurement error; Epidemiology; Methods; Asthma

1. Introduction

An analysis of incidence is the standard method to analyze the association between baseline exposure and risk of disease in a follow-up study. Cumulative incidence among both exposed and nonexposed is calculated as number of new cases of disease among the population at risk (i.e., for chronic diseases, subjects free of disease at the beginning of follow-up). Excluding subjects with prevalent disease from the follow-up is meant to ensure that the study can assess causality truly prospectively—that is, measured exposure precedes onset of disease. When calculating incidence, it is commonly assumed that there is no error in excluding prevalent cases of disease from the follow-up.

In a follow-up study, nondifferential misclassification of disease during follow-up leads in general to bias toward null in the estimated relative risks [1]. This bias is mainly a function of specificity [1], especially for rare diseases, and in the presence of perfect specificity there is no bias [2].

Effect of misclassification of disease at baseline has rarely been considered. Recently it was suggested that nondifferential misclassification of binary disease can lead to bias either away or toward null in odds ratios for transition probabilities in a follow-up study [3]. In a follow-up study, it was shown that measurement error in continuous outcomes can lead to bias either away or toward null when the analysis is adjusted for the baseline level of the outcome [4,5]. It has also recently been shown that measurement error can severely bias estimates of incidence of asthma (unpublished data, 2003). To our knowledge, however, the effect of disease misclassification at baseline on the estimated cumulative incidence risk ratios has not previously been analyzed.

Our objective here was to explore the magnitude and determinants of bias in the observed cumulative incidence
risk ratios in a follow-up study of a dichotomous disease measured with error at baseline.

2. Theory

2.1. Effect of misclassification of disease on estimated incidence

We first illustrate the effect of misclassification of disease on the estimated cumulative incidence (Fig. 1). For added clarity, all possible 16 combinations of true and observed changes in disease status in an analysis of cumulative incidence are also shown in Table 1. Note that we consider only cumulative incidence, assume that misclassification of disease is nondifferential, and do not consider misclassification of exposure. We also assume that risk of remission of disease is independent of risk of incidence and that misclassification of disease at baseline and at follow-up are independent of each other.

In Fig. 1, the column labeled ‘Baseline’ represents the result of the baseline examination, in which about one-quarter of the subjects are tested positive for the disease and thereby excluded from the follow-up. Among those who test negative, most subjects are truly free of disease (marked with white). However, due to imperfect sensitivity of the test, some subjects who truly do have disease (marked with black) nonetheless test negative.

Only subjects testing negative for the disease at baseline are included in the follow-up. For calculating incidence, they form the denominator; subjects testing positive at follow-up (observed incident cases) form the numerator.

In the presence of disease misclassification and remission of disease, observed incident cases may have four different true disease histories (the four arrows in Fig. 1 or rows 1, 3, 5, and 7 in Table 1). First, due to imperfect specificity of the test at follow-up, subjects may be misclassified as having the disease at follow-up (false positives, combinations 3 and 7), although they do not have it. The false positives include both subjects without disease either at baseline or at follow-up (combination 3) and some subjects who had disease at baseline but no longer at follow-up (i.e., subjects whose disease remits during the follow-up) (combination 7). Second, subjects may be true positives at follow-up (combinations 1 and 5). This includes true incident cases—that is, subjects without disease at baseline, but who develop it during follow-up and also test positive for it (combination 1). However, it also includes subjects who already had the disease at baseline, but who were not excluded from the follow-up due to imperfect sensitivity of the test at baseline (combination 5).

2.2. Bias in observed incidence risk ratio

The equations for calculating observed cumulative incidence and incidence risk ratios in the presence of misclassification of disease and remission are derived in the Appendix. The variables needed to calculate observed incidence risk ratios in the presence of misclassification of disease and remission are listed in Table 2. They include sensitivities and specificities of the test for the disease at baseline and at follow-up, prevalence of disease at baseline, risk ratio of prevalence between exposed and nonexposed, incidence, remission and risk ratios of incidence and remission.

In addition, we define bias in the observed incidence risk ratio as the difference between observed (IRR_{o}) and true (IRR_{t}) incidence risk ratios:

\[
\text{Bias} = IRR_{o} - IRR_{t}
\]
3.52 and 4.43. The largest biases were observed with low
sensitivity at baseline, whereas the effect of specificity is
less strong (Fig. 2). High true prevalence of disease at
also increases the potential for bias. As could be
expected, bias was largest when true incidence was low.
Bias increased continuously with increasing difference be-
tween prevalence and incidence among the nonexposed
(data not shown). If true prevalence was smaller than true
incidence, bias was usually below 1. The effect of remis-
sion is less strong and could be seen only at quite high rates
of remission among the exposed (Fig. 2).

The direction of bias in the observed incidence risk ra-
tios due to imperfect sensitivity at baseline depends on
the relative strength of the risk ratios for prevalence and
for incidence. This can be seen by plotting bias against sen-
sitivity for different values of risk ratio for prevalence and
incidence (Fig. 3). Other variables were kept fixed. When
true risk ratios for prevalence and incidence are the same,
there is little bias. In the other situations, the observed in-
cidence risk ratio reflects the risk ratio at baseline the more
the lower is the sensitivity. Therefore, if the risk ratio for
prevalence is larger than for incidence, the bias is always
upward, and vice versa. Therefore, in the presence of mis-
classification of disease at baseline the observed incidence
risk ratios depend on the association between exposure and
disease both at baseline and during follow-up.

2.3. Hypothetical examples of incidence bias

The simulations describe the general behavior of bias in
the observed cumulative incidence risk ratios. To under-
score some important behaviors of this bias, we quantify
the amount of bias for three types of situations (Table 3).

In the first situation, the risk factor is associated with
disease risk only at baseline, but not at follow-up. This sit-
uation is very possible in, for example, follow-up studies of
adult-onset asthma, because the pathogenesis and risk fac-
tors for asthma in adulthood and in childhood appear differ-
ent [6]. This situation is also especially important, in that
follow-up studies are usually undertaken to find out if an
exposure that is associated with disease cross-sectionally
also predicts incident disease. For example, most studies
on risk factors of asthma in adulthood have been cross-
sectional, but at writing there are several on-going follow-up studies [7]. There is also substantial misclassifi-
cation in asthma; questionnaires and bronchial hyperres-
sponsiveness testing for asthma have sensitivity ranging
between 0.6 and 0.9 and specificity of 0.7 to 0.9 [8]. In
this first situation, imperfect sensitivity at baseline leads
to overestimation of the observed incidence risk ratios
(Table 3, section A).

In the second situation, there is a positive association
during follow-up, but no association at baseline, as in the
case of vigorous exercise triggering sudden death [9] or
new (e.g., occupational) exposure in adult asthma. In this
situation, the observed incidence risk ratio is biased down-
wards (Table 3, section B). In the third situation, there is

---

Table 1

<table>
<thead>
<tr>
<th>Disease at baseline</th>
<th>Disease at follow-up</th>
<th>Change in disease status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combination</td>
<td>True</td>
<td>Test</td>
</tr>
<tr>
<td>1</td>
<td>− − + +</td>
<td>Incidence</td>
</tr>
<tr>
<td>2</td>
<td>− − + −</td>
<td>Incidence</td>
</tr>
<tr>
<td>3</td>
<td>− − + +</td>
<td>No disease ever</td>
</tr>
<tr>
<td>4</td>
<td>− − − −</td>
<td>No disease ever</td>
</tr>
<tr>
<td>5</td>
<td>+ − + +</td>
<td>Disease always</td>
</tr>
<tr>
<td>6</td>
<td>+ − − +</td>
<td>Disease always</td>
</tr>
<tr>
<td>7</td>
<td>+ − − +</td>
<td>Remission</td>
</tr>
<tr>
<td>8</td>
<td>+ − − −</td>
<td>Remission</td>
</tr>
<tr>
<td>9</td>
<td>− + + +</td>
<td>Incidence</td>
</tr>
<tr>
<td>10</td>
<td>− + − +</td>
<td>Incidence</td>
</tr>
<tr>
<td>11</td>
<td>− + − +</td>
<td>No disease ever</td>
</tr>
<tr>
<td>12</td>
<td>+ + + +</td>
<td>Disease always</td>
</tr>
<tr>
<td>13</td>
<td>+ + − +</td>
<td>Disease always</td>
</tr>
<tr>
<td>14</td>
<td>+ + − −</td>
<td>Remission</td>
</tr>
<tr>
<td>15</td>
<td>+ + − −</td>
<td>Remission</td>
</tr>
</tbody>
</table>

Abbreviation: EF, excluded from follow-up.

Table 2

<table>
<thead>
<tr>
<th>Variables</th>
<th>Nonexposed</th>
<th>Exposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity, $S_e$</td>
<td>0.5–1</td>
<td></td>
</tr>
<tr>
<td>Specificity, $S_p$</td>
<td>0.5–1</td>
<td></td>
</tr>
<tr>
<td>Incidence, $I$, %</td>
<td>0.5–5</td>
<td>0.05–2.5a</td>
</tr>
<tr>
<td>Remission, $r$, %</td>
<td>0.5–5</td>
<td>0.05–2.5a</td>
</tr>
<tr>
<td>Incidence risk ratio, $IRR_t$</td>
<td>0.1–5</td>
<td></td>
</tr>
<tr>
<td>Remission risk ratio, $RIR_t$</td>
<td>0.1–5</td>
<td></td>
</tr>
</tbody>
</table>

Subscripts: o, observed; t, true; 1, baseline, 2, follow-up. See further in Appendix.

* Calculated based on values of other variables in this table.
similar association both at baseline and during follow-up, which is possibly the most common situation (e.g., blood lipids and myocardial infarction) [10]. In this situation, there is little bias (Table 3, section C).

Bias is increased if the true incidence is low, because in this situation the true incident cases constitute a smaller proportion of the observed incident cases (Table 3, section D). Also, if the true prevalence of disease is high at baseline, the bias due to imperfect sensitivity at baseline is larger (Table 3, section E). High remission among exposed subjects reduces the bias due to imperfect sensitivity at baseline (Table 3, section F).

Fig. 2. Variables affecting bias (observed minus true incidence risk ratio). Plots produced by solving equations in Appendix on 10,000 datasets generated by randomly varying the variables in the ranges given in Table 2 with uniform distribution of probability. No misclassification at follow-up.
Specificity at baseline and sensitivity at follow-up have less effect. For example, given a true prevalence of 5% and otherwise the same conditions as in Table 3, section E, changing specificity of the test at baseline from 1 to 0.7 increases the observed incidence risk ratio from 2.15 to 2.33. The observed incidence among nonexposed would increase from 2.0% to 2.5% and among exposed from 4.3% to 5.7%. The same change in sensitivity at follow-up would have no effect on the observed incidence risk ratio.

The previous discussion has assumed that sensitivity and specificity at follow-up is perfect (i.e., equals 1), which can never be reached. It is well known that misclassification of a dichotomous outcome at follow-up leads to bias toward null [1]. In a cohort study, this bias is mainly a function of specificity [1,2]. The final result of the study is a function of the bias induced by misclassification at baseline and at follow-up.

### 2.4. Example using the Spanish ECRHS study

To illustrate further, let us consider a situation where we want to analyze, in the 6-year follow-up of the Spanish ECRHS study [11], whether or not atopy (here defined as specific IgE antibodies against grass) at baseline is a risk factor of incident asthma. Based on Basagana et al. [11] and unpublished data from the same study, among subjects without atopy at baseline the prevalence of asthma would be ~4%, risk ratio at baseline would be ~5, and incidence of asthma during follow-up would be ~2% (Table 4). For simplicity, we assume no remission. According to the null hypothesis of no association, the true incidence risk ratio is one. There is little agreement on the sensitivity and specificity of the different definitions of asthma, but the estimates used in Table 4 for symptoms and for the Global Initiative for Asthma (GINA) definition [12] (i.e., symptoms with bronchial hyperresponsiveness) are based on Jenkins et al. [13] and Pekkanen and Pearce [14].

In this situation, if there is no misclassification at follow-up, the observed incidence risk ratio is 2.47, instead of the true risk ratio of 1, when using symptoms to define asthma at baseline, and 3.49 when using the more specific GINA definition [12] (Table 4). This bias away from null is strongly diluted by misclassification at follow-up, especially when using symptoms to define asthma at follow-up. The incidence of asthma is also grossly overestimated. Using the more specific GINA definition [12] of disease both at baseline and at follow-up leads to an observed incidence risk ratio of 2.00. Therefore, the best combinations appears to be a sensitive definition at baseline and a definition with high specificity at follow-up, which leads to a reasonable estimate of both incidence and incidence risk ratio.

### 3. Discussion

Our present results show that nondifferential misclassification of disease at baseline, especially imperfect sensitivity, can lead to bias away or toward null in the observed cumulative incidence risk ratios. This is in contrast to nondifferential misclassification of disease during follow-up, which in general leads to bias toward null [1]. The bias described can be substantial for disease with low incidence and high prevalence, such as asthma or myocardial infarction. The results underscore the need to select a highly sensitive test for disease at baseline to exclude all diseased subjects from the follow-up.

The incidence bias described here is important mainly in studies where the prevalence of disease at baseline is high compared to incidence of disease. For example, the prevalence of asthma in young adults ranges between 2% to 12% in the Western world [15], but incidence is less than 0.5% yearly [11]. Also, the prevalence of myocardial infarction in the adult population of the United States is 3%–7% [16], but annual incidence is below 0.5% [17]. Therefore,
the follow-up studies need to be sufficiently long to ensure a sufficient number of true incident cases.

It has earlier been shown that nondifferential misclassification of disease during follow-up in general leads to bias toward null in the estimated risk ratios [1,2]. Because this bias is mainly a function of specificity, investigators have focused on making their definitions of disease as specific as possible. For example, many cohort studies on myocardial infarction are based on disease detected at hospitals, so that a combination of high specificity and low sensitivity.

### Table 3

<p>| True | Observed |</p>
<table>
<thead>
<tr>
<th>Se1</th>
<th>p, nonexp, %</th>
<th>I0, nonexp, %</th>
<th>r, exp, %</th>
<th>PRRt</th>
<th>IRRt</th>
<th>I0, nonexp, %</th>
<th>I0, exp, %</th>
<th>IRRo</th>
<th>Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Imperfect sensitivity: association only at baseline</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 5</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1.0</td>
<td>1.0</td>
<td>1.00</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1.5</td>
<td>2.7</td>
<td>1.78</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2.5</td>
<td>5.9</td>
<td>2.35</td>
<td>1.35</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3.5</td>
<td>8.9</td>
<td>2.55</td>
<td>1.55</td>
<td></td>
</tr>
<tr>
<td>B. Imperfect sensitivity: association only at follow-up</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 5</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1.0</td>
<td>3.0</td>
<td>3.00</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1.5</td>
<td>3.5</td>
<td>2.31</td>
<td>−0.69</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2.5</td>
<td>4.5</td>
<td>1.78</td>
<td>−1.22</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3.5</td>
<td>5.5</td>
<td>1.55</td>
<td>−1.45</td>
<td></td>
</tr>
<tr>
<td>C. Imperfect sensitivity: association at baseline and at follow-up</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 5</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>1.0</td>
<td>3.0</td>
<td>3.00</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1.5</td>
<td>4.7</td>
<td>3.08</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2.5</td>
<td>7.8</td>
<td>3.10</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3.5</td>
<td>10.8</td>
<td>3.07</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>D. Effect of true incidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>5</td>
<td>0.1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1.1</td>
<td>3.5</td>
<td>3.07</td>
<td>2.07</td>
</tr>
<tr>
<td>0.8</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2.0</td>
<td>4.3</td>
<td>2.15</td>
<td>1.15</td>
</tr>
<tr>
<td>0.8</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>6.0</td>
<td>8.2</td>
<td>1.37</td>
<td>0.37</td>
</tr>
<tr>
<td>0.8</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>10.9</td>
<td>13.0</td>
<td>1.19</td>
<td>0.19</td>
</tr>
<tr>
<td>E. Effect of prevalence at baseline</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1.2</td>
<td>1.6</td>
<td>1.34</td>
<td>0.34</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2.0</td>
<td>4.3</td>
<td>2.15</td>
<td>1.15</td>
</tr>
<tr>
<td>0.8</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3.1</td>
<td>8.7</td>
<td>2.79</td>
<td>1.79</td>
</tr>
<tr>
<td>0.8</td>
<td>15</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>4.3</td>
<td>14.8</td>
<td>3.41</td>
<td>2.41</td>
</tr>
<tr>
<td>F. Effect of remission among exposed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>5</td>
<td>1</td>
<td>0.1</td>
<td>3</td>
<td>1</td>
<td>2.0</td>
<td>4.4</td>
<td>2.16</td>
<td>1.16</td>
</tr>
<tr>
<td>0.8</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2.0</td>
<td>4.3</td>
<td>2.15</td>
<td>1.15</td>
</tr>
<tr>
<td>0.8</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2.0</td>
<td>4.2</td>
<td>2.08</td>
<td>1.08</td>
</tr>
<tr>
<td>0.8</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2.0</td>
<td>3.9</td>
<td>1.91</td>
<td>0.91</td>
</tr>
</tbody>
</table>

The true variables (with notation as used in the Appendix) are sensitivity (Se1) at baseline, incidence (It) and prevalence at baseline (p) among nonexposed, remission (r) among exposed, and ratios of prevalence at baseline (prevalence risk ratio, PRRt) and of incidence (incidence risk ratio, IRRt). In all situations, perfect specificity at baseline (Sp1 = 1) and remission among nonexposed 1%. No misclassification at follow-up (Se2 = 1 and Sp2 = 1).

### Table 4

Estimated effect of misclassification of asthma at baseline (with sensitivity Se1 and specificity Sp1) only or also at follow-up (with sensitivity Se2 and specificity Sp2) on observed incidence risk ratio for atopy.

<p>| True | Observed |</p>
<table>
<thead>
<tr>
<th>Se1</th>
<th>Sp1</th>
<th>p, non-atopic, %</th>
<th>I0, non-atopic, %</th>
<th>Se2</th>
<th>Sp2</th>
<th>PRRt</th>
<th>IRRt</th>
<th>I0, non-atopic, %</th>
<th>I0, atopic, %</th>
<th>IRRo</th>
<th>Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Misclassification only at baseline (symptoms)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.9</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>2.9</td>
<td>7.2</td>
<td>2.47</td>
<td>1.47</td>
</tr>
<tr>
<td>Misclassification only at baseline (GINA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>0.98</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>4.8</td>
<td>16.8</td>
<td>3.49</td>
<td>2.49</td>
</tr>
<tr>
<td>Symptoms at baseline and follow-up</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.9</td>
<td>4</td>
<td>2</td>
<td>0.8</td>
<td>0.9</td>
<td>5</td>
<td>1</td>
<td>12.0</td>
<td>15.0</td>
<td>1.25</td>
<td>0.25</td>
</tr>
<tr>
<td>Symptoms at baseline, GINA at follow-up</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.9</td>
<td>4</td>
<td>2</td>
<td>0.3</td>
<td>0.98</td>
<td>5</td>
<td>1</td>
<td>2.8</td>
<td>4.0</td>
<td>1.42</td>
<td>0.42</td>
</tr>
<tr>
<td>0.3</td>
<td>0.98</td>
<td>4</td>
<td>2</td>
<td>0.3</td>
<td>0.98</td>
<td>5</td>
<td>1</td>
<td>3.4</td>
<td>6.7</td>
<td>2.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Variables as in Table 3. No remission.

* Based on data from the Spanish ECRHS follow-up study [11]. Operational definitions of asthma used are symptoms and the Global Initiative for Asthma (GINA) definition: symptoms with bronchial hyperresponsiveness [12] (see text).
is likely [10,18]. The same applies to studies using doctor diagnoses of asthma, due to underdiagnosis of asthma [19]. The present results underscore the need to use as highly sensitive definitions of disease at baseline, but a highly specific definition during follow-up.

The most problematic situation is when there is a strong association between the risk factor and disease at baseline and the true incidence of the disease is low. In this situation, it may be very difficult to separate the association between risk factor and disease at baseline and during follow-up using analyses of incidence. Having very high sensitivity at baseline can minimize bias. Another possibility is to try to adjust the analyses for markers of undetected disease, but this also is likely to be problematic [4–5]. It is also possible to try to correct for the misclassification [3,20,21], but as the amount of misclassification is usually unknown, a sensitivity analysis is needed. Finally, other methods to analyze change in disease status [22,23] may need to be considered.

In the present report, we have outlined only the main determinants of the incidence bias. More work is needed to consider this bias in a person-years framework and to consider also risk difference. Important assumptions in the present calculations have been that risk of remission of disease is independent of risk of incidence and that misclassification of disease at baseline and during follow-up are independent of each other. Effects of differential misclassification of disease and correlated error in exposure could also be explored.

These results show that nondifferential misclassification of disease at baseline can lead to over- or underestimation of the true incidence risk ratio, because the observed incidence risk ratio reflects both the association at baseline and at follow-up. This means that the analysis is not truly prospective. This underscores the need to carefully exclude all diseased subjects from the follow-up using sensitive tests for disease at baseline when analyzing incidence. The bias described here can be substantial for disease with low true incidence and high prevalence, such as asthma or myocardial infarction.

Acknowledgments

We thank Pia Verkasalo, MD, for helpful comments on the draft manuscript.

Appendix

All terms (except risk ratios) used in the equations below are unitless proportions or probabilities between 0 and 1, but the same equations apply also to frequencies.

A1. Calculating observed cumulative incidence in the presence of misclassification of disease and remission

In a follow-up study with a baseline and a follow-up examination, observed cumulative incidence \( I_o \) is calculated by first excluding those subjects from the calculations who test positive for disease at baseline (rows 9–16 in Table 1). Then, the proportion (or number) of subjects testing positive for the disease at follow-up (rows 1, 3, 5, and 7 in Table 1) is divided by the proportion (or number) of subjects testing negative at baseline (rows 1–8) (equation [A1]).

Those testing positive at follow-up can be either true positive, which is calculated as sensitivity at follow-up \( (Se_2) \) times the proportion of subjects with true disease at follow-up \( (TD_2) \), or false positive, which is calculated as one minus specificity at follow-up \((1 - Sp_2)\) times the proportion of subjects without true disease at follow-up \( (TN_2) \).

The denominator in equation (A1) is the sum of true and false negatives at baseline, which depends on sensitivity \( (Se_1) \) and specificity \( (Sp_1) \) at baseline and prevalence of disease at baseline \( (p) \).

Therefore, in the presence of remission and misclassification of disease both at baseline and at follow-up, the observed cumulative incidence \( I_o \) between the two examinations is

\[
I_o = \frac{Se_2 \times TD_2 + (1 - Sp_2) \times TN_2}{Sp_1 \times (1 - p) + (1 - Se_1) \times p}
\]  

(A1)

where \( Se_1 \) and \( Se_2 \) are sensitivity and \( Sp_1 \) and \( Sp_2 \) are specificity of the test for disease at baseline and at follow-up, \( p \) is proportion with disease at baseline, \( TD_2 \) is proportion of subjects with true disease at follow-up, and \( TN_2 \) is proportion of subjects without true disease at follow-up.

Those with true disease at follow-up \( (TD_2) \) have either been disease free at baseline (true incident cases, rows 1 and 2 in Table 1) or have had the disease already at baseline (rows 5 and 6). The proportion of true incident cases (rows 1 and 2) depends on true cumulative incidence during follow-up \( (I_t) \) and the proportion of subjects without disease at baseline \((1 - p)\). In the presence of misclassification at baseline, only part of the disease-free subjects are correctly labeled as being disease free; that is, we need to multiply by specificity \( (Sp_1) \) at baseline. This forms the first term in equation (A2).

The proportion of subjects with disease at both at baseline and at follow-up—that is, those who were erroneously included in the follow-up and whose disease did not remit during the follow-up (rows 5 and 6)—depends on the proportion of true-disease subjects \( (p) \) who tested negative at baseline \((1 - Se_1)\) and on the probability of not remitting during follow-up \((1 - r)\). This forms the second term in equation (A2).

Therefore, the proportion of subjects with true disease at follow-up \( (TD_2) \) can be calculated as

\[
TD_2 = Sp_1 \times (1 - p) \times I_t + (1 - Se_1) \times p \times (1 - r)
\]  

(A2)

where \( I_t \) is probability of true cumulative incidence, \( r \) probability of true remission of the disease during follow-up, and other terms as defined above in equation (A1).
Those without true disease at follow-up (TN2) have either been disease free already at baseline (rows 3 and 4 in Table 1) or have had the disease at baseline, but the disease has remitted (rows 7 and 8). The proportion of subjects on rows 3 and 4 depends on the proportion of subjects without disease at baseline \((1 - p)\), probability of not developing disease during follow-up \((1 - I_t)\), and on specificity of the test at baseline \((\text{Sp}_1)\) to label them as nondiseased (first term in equation [A3]). Subjects with remitting disease (rows 7 and 8) had prevalent disease \((1 - \text{Se}_1)\), and had disease remission \((r)\) during follow-up (second term in equation [A3]).

Therefore, the proportion of subjects without true disease at follow-up (TN2) can be calculated as

\[
TN_2 = \text{Sp}_1 \times (1 - p) \times (1 - I_t) \times (1 - \text{Se}_1) \times p \times r
\]

\[
(A3)
\]

Inserting equations [A2] and [A3] into equation [A1] gives the full formula for calculating observed cumulative incidence \((I_o)\) between the two examinations in the presence of remission and misclassification of disease at baseline and at follow-up:

\[
I_o = \frac{\text{Se}_2 \times [\text{Sp}_1 \times (1 - p) \times I_t + (1 - \text{Se}_1) \times p \times (1 - r)] + (1 - \text{Sp}_2) \times [\text{Sp}_1 \times (1 - p) \times (1 - I_t) + (1 - \text{Se}_1) \times p \times r]}{\text{Sp}_1 \times (1 - p) + (1 - \text{Se}_1) \times p}
\]

\[
(A4)
\]

where \(\text{Se}_1\) and \(\text{Se}_2\) are sensitivity and \(\text{Sp}_1\) and \(\text{Sp}_2\) are specificity of the test for disease at baseline and at follow-up, \(p\) is proportion with disease at baseline, \(I_t\) is probability of true cumulative incidence, and \(r\) is probability of true remission of the disease during follow-up.

A2. Calculating observed cumulative incidence risk ratios in the presence of misclassification of disease and remission

Observed cumulative incidence risk ratio \((\text{IRR}_o)\) is calculated as observed incidence among exposed \((I_{\text{E}E})\) divided by observed incidence among nonexposed \((I_o)\). That is,

\[
\text{IRR}_o = \frac{I_{\text{E}E}}{I_o}
\]

\[
(A5)
\]

Observed incidence among the nonexposed \((I_o)\) can be calculated using equation [A4], if we define \(p\), \(I_t\), and \(r\) as prevalence, true cumulative incidence, and true remission among nonexposed, respectively. Because here we are considering only nondifferential misclassification of disease, sensitivities \((\text{Se}_1, \text{Se}_2)\) and specificities \((\text{Sp}_1, \text{Sp}_2)\) are the same among nonexposed and exposed.

Equation (A4) can also be used to calculate observed incidence among exposed \((I_{\text{E}E})\), if we define and substitute into equation (A4)

\[
p_E = \text{PPR}_t \times p
\]

\[
I_{\text{E}E} = \text{IRR}_t \times I_t
\]

\[
r_E = \text{RRR}_t \times r
\]

\[
(A6) \quad (A7) \quad (A8)
\]

where \(p_E\) is prevalence of disease at baseline among exposed and \(\text{PPR}_t\) is the true risk ratio of prevalences at baseline, \(I_{\text{E}E}\) is true cumulative incidence among exposed and \(\text{IRR}_t\) is the true incidence risk ratio, and \(r_E\) is remission among exposed and \(\text{RRR}_t\) is the true risk ratio of remissions.

The final equation for calculating \(\text{IRR}_o\) is unfortunately very complicated and has no simple arithmetic solution, mainly because it involves calculations both on the absolute and on the relative scale. Its behavior was therefore explored using simulations and examples (see discussion under Theory, section 2.2ff).

\[
\text{References}
\]


