RE: "VALIDATION STUDIES USING AN ALLOYED GOLD STANDARD"

The paper by Wacholder et al. (1) and their reply to recent correspondence (2) in the Journal offer valuable insights into and appropriate warnings about the difficulty of truly adjusting for measurement errors. However, the matrix correction method may not be as sensitive to the use of an alloyed gold standard as might be inferred from examples given by Wacholder et al. in the section entitled "Two by Two Tables" (1, pp. 1251-2).

The adjustment procedure used in their examples relies on the unwarranted assumption that the sensitivity and specificity observed among nondiseased subjects only (in the validation study using the alloyed gold standard) are sufficient for the application of the matrix correction method. The apparent sensitivity and specificity among diseased subjects should have been used as well. When the correction method makes use of the apparent sensitivities and specificities for diseased subjects (both equal to 56/100, as indicated by table 1 of the original article (1, p. 1252)) and for nondiseased subjects (36/70 and 76/130, respectively), the adjusted odds ratio is equal to 1.9, underestimating the true odds ratio of 3. By contrast, the adjusted odds ratio of 1.9 is equal to the odds ratio estimate based on the alloyed gold standard.

This is not coincidental, because the matrix correction method "leads back" to the odds ratio estimate that would be expected with the measure used for validation, not only if this is a true gold standard but also if this is an alloyed gold standard (see Appendix). It is also rather intuitive that, on the technical level, it does not matter what is agreed upon as a gold standard for the application of the matrix correction method. Contrary to the statement by Wacholder et al., this result also implies that the bias in the corrected estimate does not depend on the correlation between the errors in the alloyed gold standard and in the measure of exposure used in the epidemiologic study. The bias depends only on the quality of the alloyed gold standard through the sensitivities and specificities among diseased and nondiseased subjects, respectively. In particular, corrected odds ratios would not be expected to overestimate true odds ratios when misclassification in the alloyed gold standard is nondifferential.

Under the assumption or belief that misclassification in the measured exposure used in a case-control study is nondifferential by disease status, an investigator could assume that validation among nondiseased subjects is sufficient on the grounds that similar estimates for sensitivity and specificity could be inferred to be valid for diseased subjects as well. The investigator would thus adopt the procedure used by Wacholder et al. (1). This reasoning might be correct, provided that the validation study used a true gold standard (although the investigator should seize the opportunity to check the assumptions of nondifferentiability if validation can also be performed among diseased subjects). When an alloyed gold standard is used for validation, however, this procedure would be fallacious, because nondifferential misclassification in the alloyed gold standard and in the usual measure (used in the study) do not ensure nondifferentiability of misclassification of the alloyed gold standard in the usual measure (as noted previously and according to Wacholder et al.'s table 1, which gives sensitivities and specificities equal to 36/70 and 76/130 for nondiseased subjects and 56/100 and 56/100 for diseased subjects in the validation of the usual measure against the alloyed gold standard, and despite nondifferential misclassification of true exposure in both usual measure and alloyed gold standard). Therefore, similar estimates for the sensitivity and specificity of the measure used in the study should not be expected for diseased and nondiseased subjects, and application of the matrix correction method should proceed accordingly, i.e., using
the sensitivity and specificity observed among the diseased and nondiseased, respectively.

REFERENCES

Frédéric Lagarde
Lars Alfredsson
Institute of Environmental Medicine
Karolinska Institute
S-171 77 Stockholm, Sweden

APPENDIX

Let $C_T$, $C_Z$, and $C_W$ be the vectors of cell counts in the correctly classified table, in the table based on the alloyed gold standard, and in the table based on the usual measure, respectively.

Let $M_Z$ and $M_W$ be invertible misclassification matrices corresponding to the alloyed gold standard and to the usual measure, respectively, i.e.,

$$C_Z = M_Z C_T,$$  \hspace{1cm} (A1)

$$C_W = M_W C_T$$ \hspace{1cm} (A2)

and

$$C_T = M_Z^{-1} C_Z,$$ \hspace{1cm} (A3)

$$C_T = M_W^{-1} C_W$$ \hspace{1cm} (A4)

Let $M_{WZ}$ be an invertible misclassification matrix corresponding to the misclassification of the alloyed gold standard in the usual measure, i.e.,

$$C_W = M_{WZ} C_Z.$$

Now equations A2 and A3 imply that $C_W = M_W M_{Z}^{-1} C_Z$, giving the misclassification matrix resulting from validation of the usual measure against the alloyed gold standard ($M_W M_{Z}^{-1} = M_{WZ}$).

Using, as specified by the correction method (3), the inverse of that matrix to obtain corrected cell counts gives

$$M_{WZ}^{-1} C_W = M_Z M_{W}^{-1} C_W = M_Z C_T = C_Z$$

(using equations A4 and A1), which are the cell counts that would be expected using the alloyed gold standard.

THE AUTHORS REPLY

We thank Drs. Lagarde and Alfredsson (1) for contributing to our understanding of the effects of misclassification when using an alloyed gold standard (2). They point out that the misclassification probabilities for the usual measure $W$ with respect to the alloyed gold standard $Z$ can be differential between cases and controls, even when $W$ and $Z$ are nondifferential with respect to the true gold standard $X$. The apparent sensitivity of $W$ using $Z$ as a gold standard depends on the relative frequencies of $(X = 1, Z = 1)$ and $(X = 0, Z = 1)$, because $Z = 1$ arises from $X = 0$ and $X = 1$ with fixed proportions. If the distribution of $X$ differs between cases and controls, the relative frequencies also differ, and the result will be differential misclassification of $W$ with respect to $Z$. Lagarde and Alfredsson (1) further contend that applying the misclassification matrix from the validation study in the controls to the cases when the error is differential is the source of the potential exaggeration of effect; however, using estimates of misclassification matrix for $W$ with respect to $Z$ from cases and controls separately will lead to the same estimates as if $Z$ had been used.

While the argument is essentially correct, one of the authors' claims and the penultimate paragraph of their Appendix (1, p. 1176) may be misleading. When nondifferential misclassification is assumed, as is the usual practice, the bias in the corrected estimate does depend on the correlation between the errors in $Z$ and $W$, as demonstrated clearly in Table 5 of our original paper (2, p. 1254). Secondly, in their Appendix, the matrix $M_{WZ}$ is consistent with the same $M_{WZ}$ and $M_{ZX}$; for example, if $X = 1$ and $X = 0$ each have the probability 0.5 and $W$ and $Z$ each have a sensitivity and specificity of 0.6 for $X$ in cases and in controls, there are infinitely many possibilities for $M_{WZ}$; these include, when $Z$ equals $W$, $M_{WZ}$ being the identity matrix (100 percent sensitivity and specificity), and, when the errors in $Z$ and $W$ are independent, $M_{WZ}$ being the matrix with entries $0.52 = 0.6^2 + 0.4^2$ on the diagonals and $0.48 = 2 	imes (0.6 	imes 0.4)$ for the nonzero off-diagonals. Notwithstanding this point, Lagarde and Alfredsson are correct in saying that, in expectation, the odds ratio estimate obtained after correction of $W$ with respect to $Z$ differentially equals what would be observed if $Z$ had been used throughout, regardless of the specific $M_{WZ}$.

We continue to be wary of correction methods based on validation studies using alloyed gold standards that are far from the truth. The cost of the validation study may become prohibitive when data from cases and controls must be used separately (3). However, when feasible, allowing for differential error may be more robust than assuming nondifferentiability, and therefore is worth considering if one plans to use a validation study to correct for misclassification.

REFERENCES