Meta-analysis-"Does one bad apple spoil the barrel? Letter to the editor

Ankum, W.M.; Mol, B.W.J.; van der Veen, F.; Bossuyt, P.M.M.

Publication date
1997

Published in
Fertility and Sterility

correspondents and to calculate the mean differences between the errors by 3D and 2D ultrasound. The mean difference is 0.78 with a standard error of 0.156 giving $T = 4.98$, $P < 0.0001$. *The editor might add that for those situations where supplemental statistical measures are necessary to separate the variance component due to measurement error from that due to the variability of the items being measured (follicle sizes), the omnibus test by Blackwood and Bradley would be appropriate (5). This approach provides a simultaneous test of the equivalence of the means and variances. The Blackwood–Bradley approach is more appropriate for the analysis of data in this study because it is designed to test for equality of variances in correlated samples compared with the more frequent situation where the study samples are independent. Identifying the variance component of measurement error and testing for equality of variance should satisfy the concerns of the correspondents with respect to the superior precision of 3D versus 2D ultrasonography.*

Measurement error is one of the most important considerations in clinical medicine, but the effect of measurement error on clinical practice continues to be underestimated. The procedures for exploring agreement between measurements should still be based on the nature of the study, but it is unfortunate that the levels of agreement approach is not widely promulgated among clinical investigators, editors, statisticians, and journal referees in the U.S. Levels of agreement analysis is a better index of measurement error for the clinician because it allows a direct interpretation that can be applied to individual measurements. The architects of this approach (Bland and Altman) received only finite unveiling in U.S. journals to date. With the consolidation of the publishing industry, perhaps their popular column, "Statistics Notes," that is published in the British Medical Journal could be syndicated and made available to a wide range of clinical journals.

Paul G. McDonough, M.D., Editor, Letters

REFERENCES

1. Kyei-Mensah A, Zaidi J, Pittrof R, Shaker A, Campbell S, Tan SL. Transvaginal three dimensional ultrasound: accuracy of follicular volume measurements. Fertil Steril 1996;65:371–6.
2. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;I:307–10.
3. Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich JA, et al. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 1985;230:1350–4.
4. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. Lancet 1995;346:1085–7.
5. Blackwood LG, Bradley EL. An omnibus test for comparing two measuring devices. J Quality Technol 1991;23:12–6.

Meta-Analysis—"Does One Bad Apple Spoil the Barrel?"

To the Editor:

Ankum et al. (1) recently described a meta-analysis of etiological studies addressing ectopic pregnancy (EP). They observe that there has been no quantitative summary of the known risk factors. Unfortunately, they did not ask "Why?" and performed the "first" meta-analysis for ectopic pregnancy (EP). Their results, partly erroneous and reflecting an outdated knowledge of EP risk factors, are unsatisfactory. This is not surprising.

Meta-analysis (a statistical summary of numerical outcomes of several studies) has been used uncritically to [1] increase statistical power for endpoints, [2] reconcile conflicting studies, and [3] improve estimates of effect size. Some governments considered meta-analysis a cheap source of scientific evidence, resulting in its becoming a "cornucopia of fundings for grants" (2). However, meta-analysis has many pitfalls (3), especially for observational studies, as fully illustrated by Ankum et al.

1. Inclusiveness of meta-analysis, and quality of included studies. The meta-analysis process for identification of papers (screening medical databases) selects all studies, however bad. Evidence should be weighted according to its quality (instruments for measuring the methodological quality of a study are available). This was not done by Ankum et al. and evidence from poor studies overwhelmed that from good studies. Also, Ankum et al. missed several important papers (fulfilling their inclusion criteria), particularly those addressing such risk factors as smoking or induced conception (4, 5).

2. Adding apples and oranges, a crucial problem for meta-analysis. Different studies often have different endpoints and risk factor definitions. The study by Ankum et al. reveals this problem for pelvic inflammatory disease (clinical or laparoscopical evidence?), pelvic/abdominal surgery (why combine them?), vaginal douching (how?). The consequence was heterogeneous results for many risk factors (even the powerless statistical tests were significant), rendering them uninterpretable.

3. Effects are often multivariate rather than univariate in epidemiology and especially for EP, with its many related risk factors (4). Most studies used by Ankum et al. struggled to identify and to adjust for confounding factors, who then comically struggled to find and mix crude odds ratios!

4. Other problems illustrated by Ankum et al. are nonlinear relationships between some risk factors and EP risk (for example, smoking) and effects of grouping due to controls of different nature. Eysenck (3) quoted Rutherford, ". . . when you need statistics to make your results significant, you would be better off doing a better experiment." EP epidemiology needs well-de-

signed studies addressing pertinent hypotheses (5), not meta-analyses.

Joël Coste, M.D., Ph.D.
Jean Bouyer, Ph.D.
Nadine Job-Spira, M.D.
INSERM U292
Le Kremlin-Bicêtre Cedex, France
August 12, 1996

REFERENCES

1. Ankum WM, Mol BWJ, van der Veen F, Bossuyt PMM. Risk factors for ectopic pregnancy: a meta-analysis. Fertil Steril 1996;65:1093–9.
2. Shapiro S. Meta-analysis/Shmeta-analysis. Am J Epidemiol 1994;140:771–8.
3. Eysenck HJ. Meta-analysis and its problems. Br Med J 1994;309:789–92.
4. Coste J, Job-Spira N, Fernandez H, Papiernik E, Spira A. Risk factors for ectopic pregnancy: a case control study in France, with special focus on infectious factors. Am J Epidemiol 1991;133:839–49.
5. Fernandez H, Bouyer J, Coste J, Job-Spira N. The hidden side of ectopic pregnancy: the hormonal factor. Hum Reprod 1996;243–4.

Reply of the Authors:

Coste et al. have made critical comments on our meta-analysis on risk factors for ectopic pregnancy (EP). We disagree with their statements that our results are partly erroneous and reflecting an outdated knowledge.

Our meta-analysis focused on studies using pregnant and nonpregnant controls. Studies using pregnant controls provide knowledge on the risk for EP in early pregnancy. This is of clinical importance in the diagnosis of EP, for instance, in women undergoing screening. Studies using nonpregnant controls are of less clinical importance, but they generate knowledge on the etiology of EP. Recently delivered women are less appropriate as a control group since the presence of risk indicators is decreased compared with pregnant controls. Studies using delivered controls, therefore, are likely to generate inflated odds ratios.

The paper by Coste et al. was in this category. It was not missed by our search strategy, but simply excluded from further analysis because it used the wrong control group. The paper by Fernandez et al. was published well after submission of our paper, and it seems to report on data from the same study in which delivered controls were used.

We agree with Coste et al. when they state that in meta-analysis, evidence should be weighted for its quality. For example, the Cochrane Collaboration only allows evidence from randomized controlled trials to be included in their database. Unfortunately, there is no unanimity as to how the quality of studies in meta-analysis—especially for observational studies—should be quantified. We also struggled with this problem, but decided not to weigh quality issues. Instead, we excluded studies with less appropriate methodology from our analysis, e.g., studies using delivered controls. Studies may show conflicting results due to sampling variation or due to differences in design or in control definition. One should not pool the results from such studies, a principle strictly adhered to in our analysis. Whenever "apples and oranges" were discovered, we reported only the odds ratios from the original studies without giving a common odds ratio (Chlamydia titers, pelvic/abdominal surgery, medical and spontaneous abortion, infertility, tubal pathology, and vaginal douching).

As stated by Coste et al., many of the papers used in the meta-analysis struggled to identify and adjust for confounding factors. Obviously, these efforts are important when unraveling the etiology of ectopic pregnancy: theoretically, multivariate analysis of confounding factors is possible in a meta-analysis whenever sufficient crude data from the original studies are available. As stated in our paper, this was not the case in the majority of studies. Because we did not intend to unravel the etiology of EP, we were satisfied with crude odds ratios, which served our purpose quite sufficiently.

Meta-analysis is a well-formalized method to summarize available evidence on a subject that cannot be performed without well-designed studies. The discontent of our French colleagues with meta-analysis, in general, is hard to understand. The fact that the method is advocated by governments should not be a reason to disqualify its use.

*Willem M. Ankum, M.D., Ph.D.**
Ben W. J. Mol, M.D.†
*Fulco Van der Veen, M.D., Ph.D.**
Patrick M. M. Bossuyt, Ph.D.†
*Departments of *Obstetrics and Gynecology (H4-205)*
and †Clinical Epidemiology and Biostatistics
Academic Medical Center
University of Amsterdam
Amsterdam, The Netherlands
December 11, 1996