Journal of Clinical Epidemiology 66 (2013) 355–358

Journal of Clinical Epidemiology

COMMENTARIES

The withholding of test results as a means of assessing the effectiveness of treatment in test-positive persons

Noel S. Weiss*

University of Washington, Box 357236, 1959 NE Pacific St. Room F262D, Seattle, WA 98195, USA

Accepted 27 July 2012; Published online 13 September 2012

Abstract

In recent years, a number of studies have achieved randomization of patients to alternative management strategies by blinding some patients (and their providers of medical care) to the results of tests that guide such strategies. Although this research approach has the potential to be a powerful means of measuring treatment effectiveness, the interpretation of the results may not be straightforward if the treatment received by test-positive persons is variable or not well documented, or if the analysis is not restricted to outcomes in test-positive persons. Studies in which the test results are withheld at random may face ethical issues that, to date, have received little discussion.

Keywords: Randomized trials; Treatment efficacy; Medical ethics

The influence of a given treatment is most commonly assessed through a comparison of outcomes in persons who receive and do not receive that treatment. For example, in the late 1960s, because the impact of blood pressure lowering was unknown, randomized trials of antihypertensive therapy were conducted among persons found to have a blood pressure elevation on a screening examination. Less commonly, some investigators have sought to address the effectiveness of the test-prompted treatment in another way, by performing a diagnostic or screening test on all the potential participants but informing just some patients (and their providers) of their test results. In such a study, test-positive persons who are apprised of their status may receive active therapy; those who test positive and are kept in the dark (and whose providers are kept in the dark) receive no treatment at that time, although they may do so if and when clinically evident manifestations of their condition develop later on.

The goal of this commentary is to provide examples of studies of the effectiveness of treatments administered as a result of findings on a diagnostic or screening test in which positive test results were provided only to some patients in the study population. It also discusses how the analysis of the results of such studies needs to be approached to maximize the potential of this design.

Example 1. A total of 1,442 women with ovarian cancer in clinical remission after receipt of the platinum-based chemotherapy were followed every 3 months with a clinical examination and assessment of serum levels of CA125 [1]. Women whose CA125 value was more than twice the average of those without cancer were randomly assigned to receive or not receive additional chemotherapy at that time (265 and 264 women, respectively). Women in the latter (delayed) group were not informed regarding the positive test results nor were their physicians so informed. The presence of clinical recurrence at a later time in a woman in either group could give rise to a receipt of further chemotherapy (88% of women in the delayed group ultimately received second-line chemotherapy).

Example 2. In association with a visit to their physician, more than 6,000 women were asked to complete a questionnaire regarding the intimate partner violence but assigned at random to do so either before or after that visit [2]. A total of 347 women who answered the questions before the visit met the study criteria for intimate partner violence and were provided with an information card regarding resources for women experiencing violence. Their physicians were informed of the questionnaire information and carried out

* Corresponding author. Tel.: 206-685-1788; fax: 206-543-8525.
E-mail address: nweiss@uw.edu

0895-4356/$ - see front matter
http://dx.doi.org/10.1016/j.jclinepi.2012.07.011

a discussion with and/or made referrals for each patient as deemed appropriate. These women were requeried 18 months later, along with the 360 women who on the post-visit completion of the questionnaire met the criteria for intimate partner violence and who received only the information card, for a history of the recurrence of intimate partner violence and quality of life during those 18 months.

Example 3. Lazarus et al. [3] examined serum specimens of a large number of women early in pregnancy for indications of hypothyroidism. In a randomly selected half of the women, the specimens were tested immediately, and women whose free T_4 level was below the 2.5th percentile (or whose thyrotrophin level was above the 97.5th percentile) were given thyroid hormone treatment for the duration of the pregnancy. In the other women, the serum specimens were stored and tested at the end of the pregnancy. At 3 years of age, the children of those women in both groups whose initial serum sample met the biochemical criterion for hypothyroidism were compared for cognitive function. The lack of a difference in the cognitive function of these children suggested a lack of efficacy (in this dimension) of thyroid hormone treatment of pregnant women with mild hypothyroidism.

Example 4. Among patients with suspected acute coronary syndrome seen during February 2008 through July 2009 in a Scottish dispensary, plasma troponin levels were routinely assessed [4]. Although during the first 6 months the threshold level for action was recommended to be $\geq$0.20 ng/mL, during the last six months, it was $\geq$0.05 ng/mL: values below these thresholds were not reported to clinicians. During a 1-year follow-up period, the investigators documented the incidence of death and recurrent myocardial infarction among persons with plasma troponin levels between 0.05 and 0.19 ng/mL, comparing those seen in the earlier period (who, on average, were not managed aggressively) and those seen in the later period (who generally received more aggressive therapy).

Whether randomization is used (Examples 1−3) or not (Example 4), the value of this design stems from its ability to focus its analysis on outcomes in the patients who test positive, and becomes essentially an evaluation of the effectiveness of therapy in such persons. Without the knowledge of test results in both groups of patients, an evaluation of the impact of treatment would be indirect, necessitating a comparison of outcomes between all tested and nontested persons in the study population. To the extent that the outcome(s) of concern may occur both in persons who do and do not test positive, any beneficial influence of treatment will be diluted. For example, consider women in an initial remission of advanced ovarian cancer who later die of their disease without having had an elevation of serum CA125 before the recurrence of their disease. It would be necessary to include such fatalities in the comparison of mortality between women who are and are not tested for CA125 levels, but the expected equal number of these deaths in the two groups would act to lead to an underestimation of any true reduction in mortality associated with the receipt of treatment in test-positive women.

The following is an example of a trial in which, although the results of testing were made available to providers of half of the study participants, the analysis was not confined to the outcomes in test-positive persons.

Example 5. HIV-infected adults with a CD4 count less than 200 cells/μL seen in four centers in sub-Saharan Africa were started on antiretroviral therapy [5]. These patients were asked to return to see a study physician every 12 weeks; at that time, a blood sample was obtained and a red and white cell counts were performed (including a CD4 cell count), in addition to biochemical indicators of liver and kidney functions. The patients were assigned at random to:

a) have their physician be routinely notified of all laboratory results or
b) have their physician not be notified, unless he/she requested results because of clinical indications. However, the CD4 count was not provided even if requested.

During a follow-up of about 5 years, on average, the incidence of death and disease progression was compared between these two groups.

The ratio of the 5-year cumulative mortality between the two groups—10% in the patients whose laboratory results were routinely made available and 13% in the other patients—is undoubtedly an underestimate of the relative benefit of treatment among patients in whom a laboratory abnormality (e.g., low CD4 count) occurred during the follow-up. About half the deaths in study participants occurred in those whose last CD4 count exceeded 100 cells/μL. Because a criterion for switching to a second-line agent was a CD4 count of 100 or lower, about half the deaths tabulated in the analysis occurred among persons who would not have benefited from a modification of the treatment regimen stimulated by the laboratory monitoring program. A more sensitive analysis would have restricted attention to just those patients in both groups whose CD4 count fell below 100 (or who had some other laboratory abnormality that would have necessitated treatment).

The final example of a study in which the knowledge of test positivity is withheld illustrates the magnitude of the impact of failing to confine the analysis to persons with a positive test result.

Example 6. Vaginal swabs were obtained from 2,529 sexually active women aged 16−27 years [6]. The swabs were tested for the presence of *Chlamydia trachomatis*, right away in 50% of the women and not until 12 months later (after having been frozen at −80°C) in the others.