ELSEVIER

Journal of Clinical Epidemiology 66 (2013) 202−208

Journal of Clinical Epidemiology

Matching of controls may lead to biased estimates of specificity in the evaluation of cancer screening tests

Hermann Brenner[a,*], Lutz Altenhofen[b], Sha Tao[a]

[a]*Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, INF 581, D-69120 Heidelberg, Germany*
[b]*Central Research Institute of Ambulatory Health Care in Germany, Hervert-Lewin-Platz 3, D-10623 Berlin, Germany*

Accepted 24 September 2012

Abstract

Objectives: In the evaluation of cancer screening tests, cancer-free controls are often matched to cancer cases on factors such as sex and age. We assessed the potential merits and pitfalls of such matching using an example from colorectal cancer (CRC) screening.

Study Design and Setting: We compared sex and age distribution of CRC cases and cancer-free people undergoing screening colonoscopy in Germany in 2006 and 2007. We assessed specificity by sex and age of two immunochemical fecal occult blood tests (iFOBTs) in a study among screening colonoscopy participants conducted in the same years, and we assessed the expected impact of matching by sex and age on the validity of specificity estimates at various cut points.

Results: In the screening colonoscopy program, the proportion of men and mean age were 59.6% and 68.6 years among 10,324 CRC patients compared with 45.6% and 64.7 years, respectively, among 997,490 cancer-free participants. The specificity of the iFOBTs was higher among women than among men and decreased with age. Matching of cancer-free controls by age and sex would have led to the underestimation of specificity at all cut points assessed.

Conclusion: In the evaluation of cancer screening tests, matching of controls may lead to biased estimates of specificity.

Keywords: Bias; Early detection; Matching; Screening; Specificity; Statistical methods

1. Introduction

In the ‘‘omics’’ era, research on novel early detection markers for cancer is blooming [1,2]. In the evaluation of cancer early detection markers, cancer patients are recruited for determining sensitivity. For the evaluation of specificity, noncancer controls are needed. Ideally, cancer-free controls should be recruited from the screening population, but, given the difficulty of verification of the absence of cancer by a (often rather invasive) gold standard method in screening populations, cancer-free convenience samples recruited in clinical settings are often used in practice. The suitability of controls is often judged by their comparability with cases with respect to the distribution of key sociodemographic factors, such as sex and age. In some studies, matching by these factors is used to ensure full comparability. For example, in a recent systematic review on the performance of blood-based tests for early detection of colorectal cancer (CRC) [3], information on sex and age of cases and controls was given in approximately half of the studies. While many studies used convenience samples of controls (such as patients with benign diseases or blood donors) that were on average considerably younger than cases, matching of controls by age or sex and age was explicitly reported in four studies [4−7].

However, for valid judgment of the specificity of cancer early detection markers, the controls should be representative of cancer-free people from the screening population, who might differ from the cases with respect to sex and age. For example, because of the strong rise of both cancer incidence and prevalence with age [8,9], cases would almost always be expected to be on average older than noncases among screening participants. Likewise, the strong

Conflict of interest statement: None.

Competing interests: There are no competing interests.

Grant support: The German screening colonoscopy registry is funded by the National Association of Statutory Health Insurance Physicians and the National Health Insurance Fund. The BLITZ study was supported in part by the German Research Foundation (Deutsche Forschungsgemeinschaft) within the framework of a PhD program (Graduiertenkolleg 793) and by a grant from the German Federal Ministry of Education and Research (O1ESO72). The test kits were provided free of charge by the manufacturer. The sponsor did not have any role in this study.

* Corresponding author. Tel.: +49-6221-421301; fax: +49-6221-421302.

E-mail address: h.brenner@dkfz.de (H. Brenner).

0895-4356/$ - see front matter
http://dx.doi.org/10.1016/j.jclinepi.2012.09.008

What is new?

- Matching of controls to the sex and age distribution of cancer cases is commonly used, and the degree of matching is often perceived to be a quality criterion in studies evaluating cancer early detection tests.
- The authors show that, in contrast to widespread belief, ''perfect matching'' may in fact lead to biased estimates of specificity, and they illustrate the potential merits and pitfalls of matching using the example of colorectal cancer screening studies.

sex differences in incidence and prevalence seen for many forms of cancer, such as lung, stomach, and CRC or skin melanoma, would be expected to result in major differences in the sex distribution of cases and cancer-free controls [8,9]. Unsurprisingly, age and sex were found to be strongly associated with findings of cancer and precancerous lesions in many cancer screening studies [10−12].

Implications, pros, and cons of matching as a tool to prevent confounding or enhance precision and power have been widely addressed in the context of assessing risk factor effects in cohort and case−control studies, respectively [13−15]. To our knowledge, no previous article has addressed the implications of matching in studies aiming to assess the performance characteristics of screening tests. In this article, we aim to assess potential implications of such matching. We illustrate by an empirical example from the setting of CRC screening that matching of controls to the sex and age distribution of cases might lead to biased estimates of specificity if the sex and age distribution of people in the screening population vary between those with and without the disease and specificity likewise varies according to sex and age. Our illustration is based on data from the German national screening colonoscopy program [16,17] and from a study evaluating two immunochemical fecal occult blood tests (iFOBTs) among participants of this program [18,19].

2. Methods

2.1. Databases

Data from the German national screening colonoscopy registry were used to assess the sex and age distribution of CRC cases and cancer-free participants of screening colonoscopy. Colonoscopy is the current gold standard for diagnosis of CRC. In Germany, screening colonoscopy is offered as a primary screening examination for early detection and prevention of CRC since October 2002. Women and men are eligible for a first screening colonoscopy from the age of 55 years. If this first screening colonoscopy is conducted before 65 years of age, a second screening examination will be offered 10 years later. Certification to conduct screening colonoscopy is tightly regulated on the basis of extensive previous training and experience, and maintenance of certification is contingent on conducting at least 200 colonoscopies and 10 polypectomies per year that are subject to rigorous quality control. Histopathologic examination is performed decentrally by certified pathology laboratories.

Details on the national German screening colonoscopy registry have previously been reported [16,17]. Briefly, the results of all screening colonoscopies are reported on a standardized form by the physicians. Reporting is considered virtually complete as it is a prerequisite for reimbursement for colonoscopies by the health insurance funds. The registry includes only colonoscopies conducted as primary screening examinations (i.e., colonoscopies conducted for work-up of results from other tests, such as positive results of fecal occult blood tests (FOBTs), because of symptoms or for surveillance of previous findings, which are separately reimbursed as ''therapeutic colonoscopies'' are not included). Items reported include basic sociodemographic variables and information on findings at colonoscopy. The reporting forms are scanned, processed, and checked for completeness and plausibility using standardized algorithms at regional data centers before anonymized transfer to the national data center is performed. Approximately 3% of eligible people participate in screening colonoscopy each year, which translates to an expected participation rate of 25−30% during the 10-year time window foreseen for this screening option. For this analysis, we used data from 1,007,814 first-time screening colonoscopies in 2006 and 2007. This time window was chosen as it corresponds to the time window of data collection for the early detection marker evaluation study described in the following.

iFOBTs are increasingly recommended and used for CRC screening because of advantages in test performance and acceptance over traditional guaiac-based FOBTs [20−24]. Estimates of the specificity of two iFOBTs were derived from the BLITZ study, a study among participants of screening colonoscopy in Southern Germany which has been described in detail elsewhere [18,19,25−27]. Briefly, 1,785 participants were recruited in 20 gastroenterology practices between January 2006 and December 2007 according to a protocol approved by the ethics committees of the Medical Faculty Heidelberg of the University of Heidelberg and physicians' chambers of Baden-Württemberg, Rheinland-Pfalz, and Hessen. Patients were informed about the study at a preparatory visit in the practice, typically about 1 week before colonoscopy. They were asked to provide a stool sample before bowel preparation for colonoscopy, which was used for the evaluation of multiple stool-based early detection markers, including multiple qualitative [18,25] and quantitative iFOBTs [19,26,27] in a central laboratory. Results for two quantitative tests (RIDASCREEN Haemoglobin and RIDASCREEN Haemo-/Haptoglobin Complex;