Identification of physicians with unusual performance in screening colonoscopy databases: a Bayesian approach

Christian Stock, PhD,[1] **Lorenz Uhlmann, MSc,**[1] **Michael Hoffmeister, PhD,**[2] **Gunter Laux, PhD,**[3] **Meinhard Kieser, PhD,**[1] **Hermann Brenner, MD, MPH**[2]

Heidelberg, Germany

Background: The adenoma detection rate (ADR) is an important surrogate measure of performance quality for screening colonoscopies.

Objective: To demonstrate how physicians with unusual performance concerning the adenoma detection rate may be identified in screening colonoscopy databases.

Design: Bayesian random-effects modeling and Winsorization of potential outliers were applied to develop a robust model for the majority of providers. Divergence was assessed with adjustment for multiple testing. The steps in the analysis were visualized by using funnel plots. Additionally, minimum requirements for the number of colonoscopies with 1 or more detected adenomas were derived.

Setting: Data from 422 physicians offering screening colonoscopy and participating in a quality assurance program in Bavaria, Germany, were used.

Patients: A total of 69,738 asymptomatic individuals 55 to 79 years of age.

Intervention: Screening colonoscopy.

Main Outcome Measurements: Physician-specific ADRs.

Results: The overall ADR in the sample was 26%. From an initial model, 62 physicians (15%) were identified as potential outliers. A model with normally distributed random effects was then chosen as the robust null model. Of the potential outliers, 10 (16%) were confirmed as physicians with unusual performance at a false discovery rate of 5%. For all of them, the observed ADR was lower than expected, and together they accounted for 1.4% of all included colonoscopies.

Limitations: Analysis of routine data.

Conclusion: The applied statistical approach appears suitable to identify unusual performance in screening colonoscopy databases. Its application may help to evaluate and improve the quality of colonoscopy in population-based colorectal cancer screening programs. (Gastrointest Endosc 2015;81:646-54.)

Detection and removal of colorectal adenomas are fundamental goals of screening colonoscopy. The adenoma detection rate (ADR) (ie, the rate of colonoscopies in which at least 1 adenoma is found) therefore is an important surrogate measure of performance quality for screening colonoscopies.[1] Recent studies have found an inverse association between ADR and the incidence of interval colorectal cancer.[2,3]

Evidence also suggests that there is considerable variation in ADRs among physicians and that individual colonoscopist factors are likely to influence the ADR.[4-7] As a means of quality assurance, target ADRs such as 25% or higher for men and 15% or higher for women have been suggested for screening of asymptomatic populations 50 years of age and older.[8]

A question that is related to the minimum standards for ADRs but is addressing an even more basic level of quality assurance is whether there are physicians whose ADRs deviate substantially from the bulk of all providers of screening colonoscopy. An answer to this question may be particularly useful in colorectal cancer screening programs in the absence of ADR targets. However, a general difficulty exists in the determination of the cutoff between diverging and nondiverging performance.

In terms of the ADR, it would be of interest to identify providers with particularly low ADRs in screening colonoscopy databases. If these providers were identified, reasons for low performance quality could be further investigated and assistance might be provided to increase the quality of colonoscopy for individual performance outliers.

In this article, we demonstrate how physicians with unusual ADRs in relation to the majority of other physicians may be identified in screening colonoscopy databases.

MATERIALS AND METHODS

Brief outline of approach

A Bayesian modeling framework proposed by Ohlssen et al[9] for identification of unusual performance in health care providers was adapted.

We used a simple random-effects model to select physicians with potentially unusual performance in terms of the ADR. The influence of these physicians was then downweighted in subsequent models that were fitted to find a robust null model for the majority of providers. Finally, it was statistically tested by accounting for multiple comparisons whether the potentially outlying providers comply with the chosen null model or can be confirmed as outliers.

Bayesian estimation and inference generally differ from classic frequentist methods that are mostly seen in clinical journals in treating parameters as random variables (as opposed to constants in frequentist methods).[10] The learning process in Bayesian methods works by modifying initial probability statements about parameters before observing the data to updated or posterior knowledge that combines both previous knowledge and the data at hand. It allows hypotheses to be assessed by using a collection of parameter samples from their posterior distribution. Advantages of Bayesian methods include the possibilities to fit complex hierarchical models by using Markov chain Monte Carlo (MCMC) methods and the probabilistic (more common sense) interpretation of

confidence intervals on parameters and probability values on hypotheses. Ohlssen et al[9] note that their methods could in principle also be carried out within a classical or Bayesian paradigm, but it was found to be more practical to take a Bayesian perspective.

To visualize the results at each stage of the analysis, we further used funnel plots.[11]

Data

For the analysis, data on primary screening colonoscopies conducted in Bavaria, Germany, during the year 2009 were considered. These data were drawn from the database of a large colonoscopy quality assurance program (Qualitätsmaßnahme Koloskopie) by the Bavarian Association of Statutory Health Insurance Physicians (Kassenärztliche Vereinigung Bayerns).[12] More than 80% of physicians performing colonoscopy in Bavaria were enrolled in the program and contributed their colonoscopy documentation electronically during the study period. The documentation comprised the same data also transmitted to the German screening colonoscopy registry and additional content, including patient demographic factors, indications, process quality, findings, adverse events, and diagnoses and treatments. Compulsory data entry for important variables, routine checks of the plausibility of the documentation, compulsory image documentation and histological analyses, and the physician remuneration also depending on the completeness of the documentation are likely to ensure a good data quality. Unfortunately, data on characteristics of nonparticipating physicians (~15%) and the reasons for nonparticipation are not available. It is assumed that nonparticpating physicians have low colonoscopy volumes and account only for a small fraction of the overall number of colonoscopies. Previous research by using this database investigated quality aspects of colonoscopy, the risk of colorectal neoplasia, and diagnostic characteristics of fecal occult blood tests.[12-19] Use of the data for the present evaluation was approved by the Bavarian Association of Statutory Health insurance physicians and the Ethics Committee of the Medical Faculty at the University of Heidelberg.

The included screening colonoscopies were conducted in the population 55 to 79 years of age (70,042 colonoscopies, 459 physicians). Primary screening colonoscopy is offered in Germany for average-risk individuals 55 years of age and older (for details on the German colorectal cancer screening program, see Pox et al[20]). We excluded data from physicians contributing fewer than 20 colonoscopies (304 colonoscopies, 37 physicians). The final study population then consisted of 69,738 colonoscopies performed by 422 physicians in total.

Statistical analysis

Data preprocessing. The outcome of interest for each of the physicians ($i = 1, ..., N$) was the logarithm of the odds ratio