EDM Forum
EDM Forum Community

eGEMs (Generating Evidence & Methods to improve patient outcomes)

Publish

11-4-2014

Estimating Screening Test Utilization Using Electronic Health Records Data

Rebecca A. Hubbard
Group Health Research Institute, rhubb@uw.edu

Jessica Chubak
Group Health Research Institute, chubak.j@ghc.org

Carolyn M. Rutter
Group Health Research Institute, rutter.c@ghc.org

Follow this and additional works at: http://repository.edm-forum.org/egems

Part of the Health Services Research Commons

Recommended Citation
Hubbard, Rebecca A.; Chubak, Jessica; and Rutter, Carolyn M. (2014) "Estimating Screening Test Utilization Using Electronic Health Records Data," *eGEMs (Generating Evidence & Methods to improve patient outcomes)*: Vol. 2: Iss. 1, Article 14.
DOI: http://dx.doi.org/10.13063/2327-9214.1109
Available at: http://repository.edm-forum.org/egems/vol2/iss1/14

This Methods Empirical Research is brought to you for free and open access by the the Publish at EDM Forum Community. It has been peer-reviewed and accepted for publication in eGEMs (Generating Evidence & Methods to improve patient outcomes).

The Electronic Data Methods (EDM) Forum is supported by the Agency for Healthcare Research and Quality (AHRQ), Grant 1U18HS022789-01. eGEMs publications do not reflect the official views of AHRQ or the United States Department of Health and Human Services.

Estimating Screening Test Utilization Using Electronic Health Records Data

Abstract

Background: Electronic health records (EHRs) are increasingly used by medical providers and offer a wide-reaching source of information on utilization of preventive services. Numerous measures used for quality assessment and public reporting are estimated based on EHR data. However, sources of error and misclassification can lead to over- or underestimation of true utilization rates. EHR-derived measures of screening test use are subject to error due to misclassification of screening and diagnostic tests. The implications of this misclassification for EHR-based screening utilization estimates have not been well explored.

Objectives: We calculated the bias in estimates of screening test utilization associated with several published EHR-based algorithms for identifying screening colonoscopies and propose two simple methods to correct this bias. We apply these corrections to obtain adjusted estimates of screening colonoscopy utilization using EHR data from Group Health, an integrated health care system in Washington state.

Findings: The bias in screening colonoscopy utilization estimates ranged from an underestimation of 3 percentage points to an overestimation of 12 percentage points across classification methods. If the operating characteristics of the classification method are known or if a statistical model that returns predicted probabilities of screening indication is applied in the population of interest, this information can be used to obtain unbiased estimates through simple corrections to the utilization rates with little loss of precision. When applied to data on colonoscopies received at Group Health, we found that an unadjusted estimate was 4 percentage points higher than our adjusted estimate.

Discussion: Error in classification of tests as screening when using EHR data to study screening utilization should be accounted for in order to eliminate bias and prevent spurious findings.

Acknowledgements

This work was supported by the National Cancer Institute of the National Institutes of Health (grant number U01CA152959).

Keywords

access, cancer, screening, standardized data collection, utilization

Disciplines

Health Services Research

Creative Commons License

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License.

This empirical research is available at EDM Forum Community: http://repository.edm-forum.org/egems/vol2/iss1/14

Estimating Screening Test Utilization Using Electronic Health Records Data

Rebecca A. Hubbard, PhD; Jessica Chubak, PhD; Carolyn M. Rutter, PhD[i]

Abstract

Background: Electronic health records (EHRs) are increasingly used by medical providers and offer a wide-reaching source of information on utilization of preventive services. Numerous measures used for quality assessment and public reporting are estimated based on EHR data. However, sources of error and misclassification can lead to over- or underestimation of true utilization rates. EHR-derived measures of screening test use are subject to error due to misclassification of screening and diagnostic tests. The implications of this misclassification for EHR-based screening utilization estimates have not been well explored.

Objectives: We calculated the bias in estimates of screening test utilization associated with several published EHR-based algorithms for identifying screening colonoscopies and propose two simple methods to correct this bias. We apply these corrections to obtain adjusted estimates of screening colonoscopy utilization using EHR data from Group Health, an integrated health care system in Washington state.

Findings: The bias in screening colonoscopy utilization estimates ranged from an underestimation of 3 percentage points to an overestimation of 12 percentage points across classification methods. If the operating characteristics of the classification method are known or if a statistical model that returns predicted probabilities of screening indication is applied in the population of interest, this information can be used to obtain unbiased estimates through simple corrections to the utilization rates with little loss of precision. When applied to data on colonoscopies received at Group Health, we found that an unadjusted estimate was 4 percentage points higher than our adjusted estimate.

Discussion: Error in classification of tests as screening when using EHR data to study screening utilization should be accounted for in order to eliminate bias and prevent spurious findings.

Introduction

Electronic health records (EHR) data, including administrative data used for billing and clinical medical records collected as part of patient care, are potential sources of information on a wide range of quality measures. These data include information on a broad patient population and facilitate research and quality assessment that is representative of care in community medical practice. EHR data will become increasingly available and valuable as adoption of electronic systems expands, a process that has been accelerated by incentives for health IT provided under the Affordable Care Act.

Rates of breast, cervical, and colorectal cancer screening are examples of measures of preventive services utilization that are commonly estimated using EHR data. For instance, cancer screening measures estimated via EHR data are included among the National Committee on Quality Assurance's Healthcare Effectiveness Data and Information Set (HEDIS) measures,[1] and have also been tied to reimbursement through incentives mandated by the Affordable Care Act and the Centers for Medicare and Medicaid Services (CMS) Hospital Outpatient Quality Reporting Program.[2] Estimating these measures using EHR data is preferable to self-report of screening test utilization due to the possibility of bias in estimates based on self-report.[3,4] Use of EHR data for evaluating screening utilization is also far more efficient than manual chart review.

Despite the potential value of EHR data for estimating screening test utilization, obtaining accurate estimates is challenging for several reasons. For instance, the National Quality Forum (NQF) colorectal cancer screening measure is an estimate of the proportion of individuals ages 50–75 years who have been screened with colonoscopy within the past 10 years, flexible sigmoidoscopy within the last 5 years, or fecal occult blood test within the past year.[5] An accurate assessment of this outcome requires reliable measures of the number of individuals in the screening-eligible population and the number of individuals receiving screening with each of the tests of interest. Identifying individuals in the screening-eligible population is challenging because individuals may have recently entered the health system population, making it difficult to determine if

[i]Group Health Research Institute

they have pre-existing diagnoses that should result in their exclusion from the eligible population. Similarly, identifying screened individuals is challenging because new enrollees may have been screened prior to enrollment in the health system. Additionally, there are no unique codes for screening as distinguished from diagnostic colonoscopy. Including diagnostic colonoscopies in estimates of screening utilization will inflate the apparent proportion of individuals who are adequately screened. This may obscure important information about access to and utilization of cancer screening services and may produce biased or spurious findings regarding disparities in screening if there is differential misclassification across patient subgroups.

The extent of the error in estimates of cancer screening utilization based on EHR data attributable to misclassification of screening indication depends on the operating characteristics of methods used to classify cancer screening tests. Several previous studies have estimated the operating characteristics of approaches to distinguishing screening and diagnostic tests using administrative EHR data.[6-8] While these studies have focused on the use of administrative EHR data, similar considerations apply to studies using data from clinical encounters, if the classification approaches in question produce measures of screening indication that have errors. In the context of breast cancer screening, several algorithms have been developed for classifying mammograms as screening or diagnostics using diagnosis and procedure codes from claims data.[6,9,10] A number of algorithms have also been developed for identifying screening colonoscopies based on EHR data with varying operating characteristics.[7,11-13] The sensitivity of these approaches ranges from 70 to 90 percent and specificity from 60 to 90 percent. A natural question arises: how might estimates of screening test utilization based on EHR data be influenced by misclassification of screening and diagnostic examinations?

In this paper we explore the implications of using EHR data with imperfect information on screening indication to estimate rates of screening test utilization. We use numerical examples and simulations focusing on screening colonoscopy to demonstrate the magnitude of bias that can be induced by using an imperfect measure of screening indication. We then demonstrate via simulation studies how to correct this bias using information on the operating characteristics of the algorithm used to assign screening indication or a predicted probability that the test was performed for screening purposes. Finally, we demonstrate the application of unadjusted and adjusted measures of screening colonoscopy utilization using EHR data from Group Health, an integrated health plan and health care delivery system in Washington state.

Numerical Examples of Bias in Estimates of Screening Test Utilization

To illustrate the bias arising from imperfect ascertainment of screening-test utilization we introduce the following notation. Let Y represent the true classification of the test with Y = 1 indicating a screening test and Y = 0 indicating a diagnostic test. We assume that Y is unobserved in EHR data but that a proxy is available, denoted Y*. For instance, in the case of colonoscopy, procedure codes do not distinguish colonoscopy performed for screening versus colonoscopy performed to evaluate signs or symptoms (i.e., a diagnostic exam). However, by applying existing algorithms to administrative data it is possible to obtain a predicted probability of screening indication that can then be dichotomized to yield an indicator that the examination was a screening test. In this example, Y* represents this dichotomized predicted probability. The operating characteristics of Y* are the sensitivity, $Se = P(Y^*=1|Y=1)$, and specificity, $Sp= P(Y^*=0|Y=0)$. In studying utilization of cancer screening tests, our objective is to estimate the proportion of the population receiving a screening test, $P(Y=1)$. If we use instead $P(Y^*=1)$ as a measure of screening utilization, the difference, $P(Y^*=1) - P(Y=1)$, is the bias in our EHR-derived estimate. We can express this bias as a function of the sensitivity and specificity of Y*, $P(Y^*=1) - P(Y=1) = Se \times P(Y=1) + (1-Sp) \times (1-P(Y=1)) - P(Y=1)$. It can be easily seen that for a test with perfect operating characteristics (sensitivity and specificity equal to 1), $P(Y^*=1)$ will be unbiased. The relative impact of sensitivity and specificity on bias will depend on $P(Y = 1)$, which is unknown in practical applications.

In Table 1, we illustrate the bias in estimates of screening-colonoscopy utilization when using three different algorithms for identifying screening examinations. Previously, self-report data have been used to estimate screening-colonoscopy utilization.[14,15] However, EHR data offer the opportunity to evaluate utilization in a broad population without the bias inherent in self-report. Motivated by estimates of lower endoscopy utilization,[15] we assume that the true proportion of individuals who have been screened with colonoscopy in the past 10 years is 58.5 percent. In this setting, bias in screening-colonoscopy utilization ranges from underestimation by 5.7 percentage points to overestimation by 11.6 percentage points.

Table 1. Estimated Screening Utilization Rates and Bias in Estimated Rates

	Sensitivity (%)	Specificity (%)	Estimated Screening Rate (%)	Bias
Colonoscopy utilization, Fisher	90	58	70.1	11.6
Colonoscopy utilization, El-Serag	70.1	71.6	52.8	-5.7
Colonoscopy utilization, Adams	88.5	90.5	55.7	-2.8

Notes: Estimated screening utilization rates and bias in estimated rates relative to a true screening rate of 58.5% using alternative approaches to identifying screening and diagnostic colonoscopies using EHR data.

Correcting Bias via Direct Adjustment or Use of Predicted Probabilities

In this section we present two methods for correcting the bias in estimates of screening utilization based on imperfect EHR-based algorithms for assigning screening indication. The first relies on the existence of validation studies that have established

the operating characteristics of the classification algorithm in the population of interest. We can use the known sensitivity and specificity of the algorithm for classifying tests as screening to correct estimates through direct algebraic manipulation of the formula for bias, providing a simple approach for obtaining unbiased utilization estimates. Specifically, a corrected formula for screening-test utilization is given by $(P(Y^*=1)+Sp-1)/(Se+Sp-1)$. Utilization is first estimated using algorithm-assigned screening indication, providing an estimate of $P(Y^*=1)$. This estimate and the estimated sensitivity and specificity of the algorithm are then substituted into the expression above to produce a bias-corrected estimate of $P(Y=1)$.

The second approach to correcting estimates of screening utilization requires an algorithm that assigns individual-level probabilities of screening indication. Some methods for identifying screening tests return a predicted probability that a test was used for screening purposes. For instance, in the context of colorectal cancer screening, a LASSO algorithm has been proposed that provides the probability that a given colonoscopy was a screening test.[8] LASSO is a statistical prediction model that combines variable selection and parameter estimation under a constraint on the size of the regression coefficients.[16] Assuming that these probabilities are well calibrated, summing the predicted probabilities of screening indication in the patient population of interest provides an unbiased estimate of the number of screening tests performed in the population. A well-calibrated probability implies that the proportion of tests with screening indication is equal to the predicted probability that a test has screening indication. For instance, among all tests with predicted probability of screening indication equal to 0.2, 20 percent will truly be screening tests if the prediction model is well calibrated. If the algorithm was developed in a different population, it may not be well calibrated and might over- or underestimate the probability of screening indication in a new population. Given a well-calibrated predicted probability, the utilization rate can be estimated without bias by taking the mean of the predicted probabilities in the population of interest.

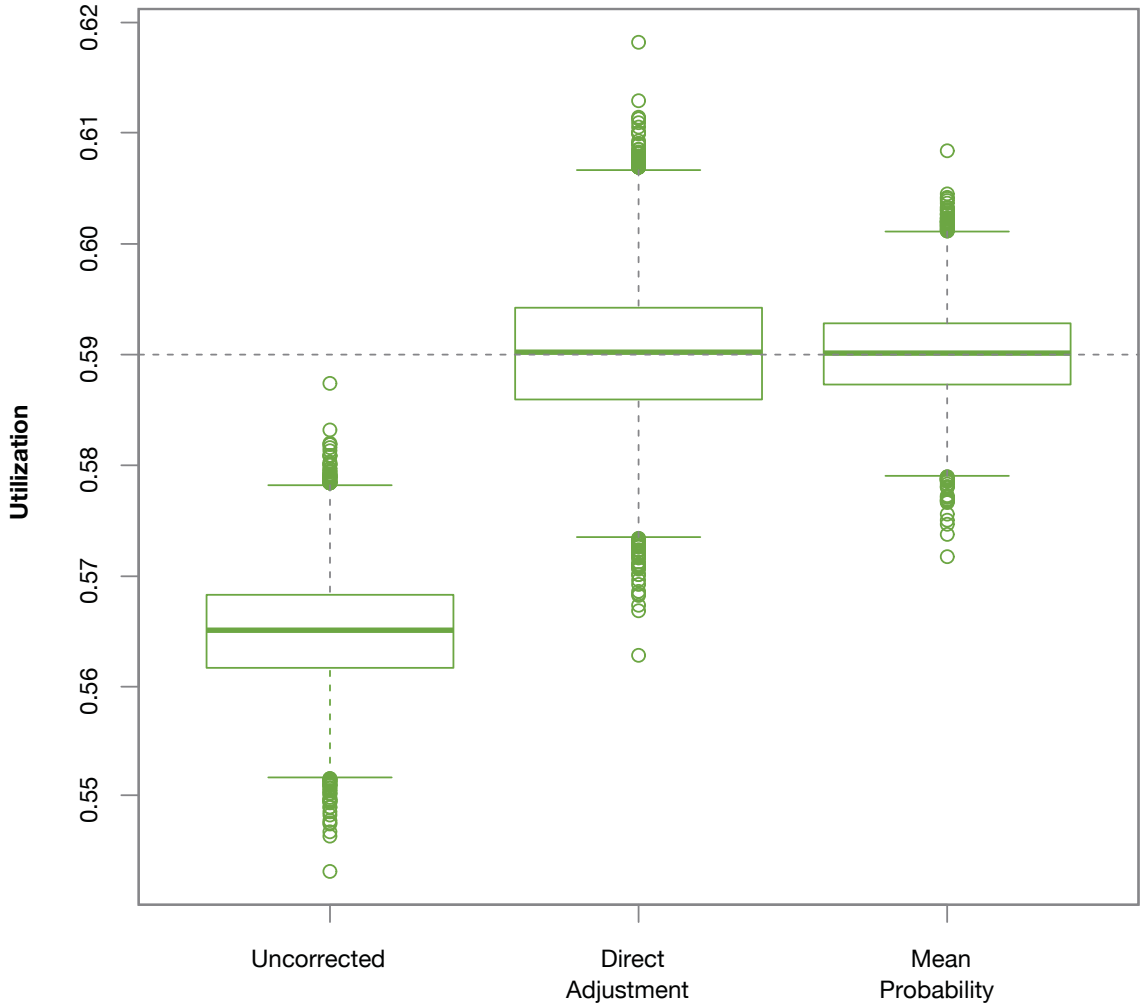
To demonstrate the bias and precision resulting from using uncorrected classifications of screening indication to estimate utilization as well as performance of the two proposed approaches to bias correction, we conducted a simulation study motivated by the context of screening with colonoscopy. In this study we assumed a population of size 10,000 and that 59 percent of this population were screened at least once over a 10-year period. We assumed that predicted probabilities of screening indication in this population arose from a Beta(0.25, 0.17) distribution. This distribution implies that the cutpoint maximizing the average of sensitivity and specificity has a sensitivity of 0.89 and a specificity of 0.90. These parameter values were selected to create a scenario similar to the Adams algorithm for classifying screening colonoscopies.[8] For each simulated individual, both the true classification of the test as screening or diagnostic and the algorithm-assigned predicted probability were known.

We first computed an uncorrected utilization estimate by dichotomizing the predicted probability at the value maximizing the average of sensitivity and specificity and then computing the proportion of individuals classified as having received a screening examination. We then computed the two bias-corrected estimates, first by applying the correction factor presented above and then by computing the mean of the predicted probabilities. We repeated this process across 10,000 simulated populations and plotted the distribution of screening-test utilization estimates provided by each approach in Figure 1. In these simulated populations, the approach directly using the algorithm-assigned classification without adjustment underestimated screening utilization by 2.5 percentage points. Both of the bias-corrected approaches had a bias of less than 0.01 percentage points. All three methods had similar precision. The uncorrected method had a standard error of 0.5, the direct adjustment method of 0.6, and the mean probability method of 0.4 percentage points.

To demonstrate the application of an adjusted utilization approach in a real data set derived from an EHR, we used data on receipt of colonoscopy from Group Health, an integrated health plan and health care delivery system in Washington state. Because sensitivity and specificity in this population are unknown we used the mean probability method to obtain adjusted estimates. We constructed a sample consisting of members ages 50 years or older who were continuously enrolled in Group Health for at least 5 years between 2002 and 2012 with no prior diagnosis of colorectal cancer. In this population, we estimated the proportion of members receiving a screening colonoscopy at least once during a 5-year period. We used Current Procedural Terminology (CPT) codes (45355, 45378–45387, 45391, 45392), Healthcare Common Procedure Coding System (HCPCS) codes (G0105, G0121) and International Classification of Diseases—Ninth Revision, Clinical Modification (ICD-9-CM) codes (45.21–45.24, 45.25, 45.42, 45.43) to identify each colonoscopy received by a member of the denominator population. We then applied the algorithm of Adams[8] to each colonoscopy to obtain a predicted probability that the colonoscopy was a screening examination. If an individual received more than one colonoscopy during this time frame we retained the colonoscopy with the highest probability of screening indication. This project was approved by the Group Health Institutional Review Board.

We first dichotomized the predicted probabilities at a threshold of 0.261, the threshold maximizing the average of sensitivity and specificity in a validation study,[8] to obtain a binary indicator of receipt of screening colonoscopy. We then computed an unadjusted measure corresponding to the proportion of individuals with a screening colonoscopy based on this binary indicator. Finally, we computed an adjusted measure by taking the mean of the predicted probabilities. This approach is appropriate for accounting for misclassification in identification of screening tests when the sensitivity and specificity of the algorithm are unknown in the target population. In this population of 139,163 individuals, the unadjusted estimate of colonoscopy screening prevalence was 12.2 percent. We compare this to an estimate of 7.8 percent based

Figure 1. Estimated Proportion of Individuals with Screening Test Utilization for 10,000 Simulated Populations



Note: The dotted line indicates true screening utilization rate.

on averaging the predicted probabilities. Thus, the unadjusted approach appears to substantially overestimate screening-colonoscopy utilization in this population.

Discussion

Estimates of screening-test utilization based on algorithms for assigning screening indication with imperfect operating characteristics will be biased. This bias leads to under- or overestimation of the proportion of the population making use of a screening test. In the case of underestimation this could lead to unnecessary efforts to improve utilization, and in the case of overestimation this could lead to failure to address the needs of underserved populations. The magnitude and direction of the bias depend on the sensitivity and specificity of the EHR-based algorithm as well as the prevalence of screening utilization in the population of interest. We have demonstrated two simple approaches to correcting bias, which can easily be applied to EHR data, and implemented an adjusted approach using EHR data from Group Health. These methods require that either the operating characteristics of the algorithm are known based on a prior validation study or that the algorithm returns predicted probabilities of screening indication and is well calibrated in the population under study. We recommend that one of these two approaches be used to correct utilization estimates whenever possible. Use of these approaches is particularly well suited to EHR data because the depth of information available via data from clinical encounters facilitates validation studies for evaluation of the discrimination and calibration of the measure in a small sample of the target population. The measure and appropriate methods for correcting for misclassification can then be rapidly applied to a broad population using administrative EHR data.

In this study, we focused on estimation of screening-test utilization when screening indication is imperfectly ascertained. However, there are a variety of other uses for EHR-derived data on cancer screening that rely on accurate ascertainment of test indications. For instance, estimates of screening-test effectiveness including the cancer stage at diagnosis and the mortality among screened and unscreened individuals may be of interest. Misclassification of screening and diagnostic tests will bias these measures of screening-test effectiveness. Several previous studies have discussed considerations for using algorithm-assigned outcomes

and exposures in studies based on EHR data.[17,18] These provide guidance on the relationship between algorithm operating characteristics and bias in various measures of effectiveness and can be used in conjunction with the current study to guide the choice of algorithm and analysis strategy when using data with known error in screening indication.

Calibration and discrimination of the EHR-based measure are key to obtaining accurate estimates. If the measure is not well calibrated all three of the approaches discussed in this paper will return biased estimates. We recommend using a validation sample, in which both clinical and administrative data are available, to evaluate the calibration of any screening indication algorithm before it is applied broadly to EHR data. Additionally, an algorithm with poor discrimination (sensitivity and specificity) will lead to biased estimates and may inflate the standard errors of the estimates. The direct adjustment method, in particular, will have inflated standard errors if its denominator, Se + Sp - 1, is close to 0. This will occur in the case of a measure where the sum of sensitivity and specificity is close to 1, which would occur, for instance, if both cases and controls are correctly classified only about 50 percent of the time.

Bias attributable to misclassification of screening indication may be particularly problematic in multisite studies. If each site uses a different approach to classifying screening indication or if a common approach is used that has differing operating characteristics across study sites, then differential bias may lead to the spurious appearance of variability in screening utilization across sites. If patient characteristics differ across sites this could also lead to the appearance of associations between patient characteristics and utilization that are entirely attributable to bias. In the case of a multisite study the classification approach should be validated at each study site allowing for bias correction at the study site level. Similarly, if operating characteristics vary across patient subgroups this could result in apparent variation in utilization across subgroups where none truly exists or, conversely, might obscure true variation in utilization. To protect against this bias, validation studies within patient subgroups of interest must be conducted to provide estimates of operating characteristics within each population or to demonstrate that algorithm-assigned probabilities are well calibrated in each group.

As use of electronic medical records expands, EHR data will become increasingly valuable as a source of information on screening test utilization. The considerations provided here allow for evaluation of the bias that will occur if algorithms for screening indication are used without correction for potential misclassification of screening tests as diagnostic tests. We have provided simple approaches to correct this bias that can be readily implemented and have demonstrated their potential for eliminating bias without inflating standard errors. The need to classify tests as screening or diagnostic is not unique to the setting of cancer screening using EHR data but exists across a broad range of prevention measures including use of tests for depression, high cholesterol, and osteoporosis. Moreover, the considerations described here do not apply solely to misclassification of diagnostic and screening tests but are relevant broadly to measures that rely on imperfect ascertainment of the service of interest. Error in ascertainment of utilization due to misclassification should be considered and appropriate adjustment methods such as those proposed in this paper should be applied to avoid bias.

Acknowledgements

This work was supported by the National Cancer Institute of the National Institutes of Health (grant number U01CA152959).

References

1. National Committee on Quality Assurance. HEDIS 2014 Volume 2: Technical Specifications for Health Plans. Washington, DC: National Committee on Quality Assurance; 2014.
2. Centers for Medicare & Medicaid Services. Hospital Outpatient Quality Reporting (OQR) Program Overview. 2014 [August 25, 2014]; Available from: http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1191255879384.
3. Cronin KA, Miglioretti DL, Krapcho M, Yu B, Geller BM, Carney PA, et al. Bias associated with self-report of prior screening mammography. Cancer Epidemiol Biomarkers Prev. 2009;18(6):1699-705. Epub 2009/06/10.
4. Rauscher GH, Johnson TP, Cho YI, Walk JA. Accuracy of self-reported cancer-screening histories: a meta-analysis. Cancer Epidemiol Biomarkers Prev. 2008;17(4):748-57. Epub 2008/04/03.
5. National Quality Forum. Quality Positioning System. [August 25, 2014]; Available from: http://www.qualityforum.org/qps/.
6. Fenton JJ, Zhu W, Balch S, Smith-Bindman R, Fishman P, Hubbard RA. Distinguishing Screening From Diagnostic Mammograms Using Medicare Claims Data. Medical care. 2012. Epub 2012/08/28.
7. Fisher DA, Grubber JM, Castor JM, Coffman CJ. Ascertainment of colonoscopy indication using administrative data. Dig Dis Sci. 2011;55(6):1721-5. Epub 2010/04/16.
8. Adams K, Johnson E, Chubak J, Doubeni C, Kamineni A, Williams A, et al. PS1-11: Development of an algorithm to classify colonoscopy indication using CRN health plan coded data. Clin Med Res 2012;10:147.
9. Smith-Bindman R, Quale C, Chu PW, Rosenberg R, Kerlikowske K. Can Medicare billing claims data be used to assess mammography utilization among women ages 65 and older? Medical care. 2006;44(5):463-70. Epub 2006/04/28.
10. Randolph WM, Mahnken JD, Goodwin JS, Freeman JL. Using Medicare data to estimate the prevalence of breast cancer screening in older women: comparison of different methods to identify screening mammograms. Health Serv Res. 2002;37(6):1643-57. Epub 2003/01/28.
11. El-Serag HB, Petersen L, Hampel H, Richardson P, Cooper G. The use of screening colonoscopy for patients cared for by the Department of Veterans Affairs. Arch Intern Med. 2006;166(20):2202-8. Epub 2006/11/15.

12. Sewitch MJ, Jiang M, Joseph L, Hilsden RJ, Bitton A. Developing model-based algorithms to identify screening colonoscopies using administrative health databases. BMC Med Inform Decis Mak. 2013;13:45. Epub 2013/04/12.
13. Ko C, Dominitz J, Neradilek M, Polissar N, Green P, Kreuter W, et al. Determination of colonoscopy indication from administrative claims data. Medical Care. 2014;52(4):e21-e9.
14. Shapiro JA, Klabunde CN, Thompson TD, Nadel MR, Seeff LC, White A. Patterns of colorectal cancer test use, including CT colonography, in the 2010 National Health Interview Survey. Cancer Epidemiol Biomarkers Prev. 2012;21(6):895-904. Epub 2012/04/12.
15. Centers for Disease Control and Prevention. Vital signs: colorectal cancer screening among adults aged 50-75 years-United States, 2008. Morbidity and mortality weekly report. 2010;59(26):808.
16. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society - Series B. 1996;58(1):267-88.
17. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol. 2005;58(4):323-37. Epub 2005/05/03.
18. Chubak J, Pocobelli G, Weiss NS. Tradeoffs between accuracy measures for electronic health care data algorithms. J Clin Epidemiol. 2012;65(3):343-9 e2. Epub 2011/12/27.