



Published in final edited form as:

Med Care. 2011 June ; 49(6): e1–e8. doi:10.1097/MLR.0b013e318207e87e.

Tradeoffs of Using Administrative Claims and Medical Records to Identify the Use of Personalized Medicine for Patients with Breast Cancer

Su-Ying Liang, PhD,

University of California, San Francisco, San Francisco, CA

Kathryn A. Phillips, PhD,

University of California, San Francisco, San Francisco, CA

Grace Wang, PhD, MPH,

University of California, San Francisco, San Francisco, CA

Carol Keohane, RN,

Brigham and Women's Hospital, Boston, MA

Joanne Armstrong, MD,

MPH Aetna, Hartford, CT

William M. Morris, BS, and

Aetna, Hartford, CT

Jennifer S. Haas, MD, MSPH

Brigham and Women's Hospital, Boston, MA

Abstract

Background—Administrative claims and medical records are important data sources to examine healthcare utilization and outcomes. Little is known about identifying personalized medicine technologies in these sources.

Objectives—To describe agreement, sensitivity, and specificity of administrative claims compared to medical records for two pairs of targeted tests and treatments for breast cancer.

Corresponding author: Dr. Su-Ying Liang, 3333 California St., Suite 420, Box 0936, University of California, San Francisco, San Francisco, CA 94143, Telephone: 415-514-0457, Fax: 415-502-0792, liangs@pharmacy.ucsf.edu.

Potential Conflicts of Interest:

Drs. Haas, Phillips, Liang and Ms Keohane received funding from a research grant from the Aetna Foundation for this research. Dr Wang has no potential conflicts of interest. Joanne Armstrong and Mike Morris are employees of Aetna. The Aetna Foundation did not have any role in the data collection, analysis and interpretation of the findings, and was not involved in manuscript approval.

Author Contributions:

Conception and Design: Armstrong, Haas, Liang, Phillips, Wang.

Acquisition of data: Armstrong, Keohane, Morris.

Analysis and Interpretation of Data: Armstrong, Haas, Keohane, Liang, Phillips, Wang.

Drafting of Manuscript: Liang.

Critical Revision of Manuscript for Important Intellectual Content: Armstrong, Haas, Keohane, Liang, Morris, Phillips, Wang.

Final Approval of the Article: Armstrong, Haas, Keohane, Liang, Morris, Phillips, Wang.

Statistical Expertise: Liang, Wang.

Obtained Funding: Haas, Phillips.

Administrative, Technical or Logistical Support: Armstrong, Haas, Keohane, Liang, Morris, Phillips, Wang.

Research Design—Retrospective analysis of medical records linked to administrative claims from a large health plan. We examined whether agreement varied by factors that facilitate tracking in claims (coding and cost) and that enhance medical record completeness (records from multiple providers).

Subjects—Women (35 – 65 years) with incident breast cancer diagnosed in 2006–2007 (n=775).

Measures—Use of human epidermal growth factor receptor 2 (HER2) and gene expression profiling (GEP) testing, trastuzumab and adjuvant chemotherapy in claims and medical records.

Results—Agreement between claims and records was substantial for GEP, trastuzumab, and chemotherapy, and lowest for HER2 tests. GEP, an expensive test with unique billing codes, had higher agreement (91.6% vs. 75.2%), sensitivity (94.9% vs. 76.7%), and specificity (90.1% vs. 29.2%) than HER2, a test without unique billing codes. Trastuzumab, a treatment with unique billing codes, had slightly higher agreement (95.1% vs. 90%) and sensitivity (98.1% vs. 87.9%) than adjuvant chemotherapy.

Conclusions—Higher agreement and specificity were associated with services that had unique billing codes and high cost. Administrative claims may be sufficient for examining services with unique billing codes. Medical records provide better data for identifying tests lacking specific codes and for research requiring detailed clinical information.

Keywords

medical record; claims data; breast neoplasm; personalized medicine

Introduction

The availability of data is critical for research designed to examine health care utilization, clinical practice patterns, and health outcomes. Administrative claims and medical record data are two common, important data sources with distinct strengths and limitations. The strengths of administrative claims data are that they are often easy to obtain for covered services with specific codes for billing. Researchers have shown that administrative data can be used to reliably identify incident cancers as well as some aspects of cancer care, such as surgeries and chemotherapy.^{1–7} One downside of administrative data is that the data reflect only the care for which payers were billed. Additionally, even for billed care, billing “bundling” may prevent administrative claims from providing adequate data about specific tests. These data may not contain pathology or radiology reports and often lack detailed clinical information.⁸

In contrast, the detailed clinical information from medical records often makes records the preferred data source for research. For example, records offer researchers information on cancer stage and comorbidities.⁸ Limitations for using medical record data are that they are costly to obtain, require abstraction, and may be incomplete.⁹ There may be additional logistical issues for getting a sample of medical records from a diversity of providers.

Many studies have attempted to assess the agreement between these two data sources to understand the utility of using one source versus the other.⁸ Studies in cancer and cancer-related care specifically have found agreement between the two data sources for use of endoscopy and chemotherapy, but these studies focus on conventional treatments and have analyzed mainly the Medicare population.^{1, 10, 11} Relatively little is known about identifying and comparing emerging technologies like personalized medicine technologies in administrative claims and medical records among a commercially insured population.

Two important examples of personalized medicine technologies are the targeted tests and treatments for breast cancer: (1) human epidermal growth factor receptor 2 (HER2) testing for trastuzumab therapy and (2) gene expression profiling (GEP) for adjuvant chemotherapy. HER2 testing for trastuzumab is one of the most successful examples of using a targeted test to determine who should receive a targeted therapy. Women whose tumors over-express HER2 may benefit from trastuzumab but women whose tumors do not over-express HER2 do not benefit from the medication. HER2 testing is recommended for all patients with invasive breast cancer,^{12, 13} and only patients with positive test results are recommended for trastuzumab treatment.¹⁴ Immunohistochemistry (IHC) and fluorescence in situ hybridization (FISH) are two tests approved to assess HER2 status. Generally, there are not unique codes to distinguish this medical indication for IHC and FISH testing, thus making the tracking of HER2 tests in claims data challenging. HER2 may also be bundled into one pathology charge and may not appear as a distinct service.

GEP is used to predict the likelihood that a patient will benefit from adjuvant chemotherapy. GEP coverage is generally restricted to patients who may benefit from testing, e.g., early-stage, estrogen-receptor-positive breast cancer patients. GEP is expensive at \$3,650 per test¹⁵ compared to <\$100 for IHC and \$300–400 for FISH¹⁶ to test HER2 status. Since 2006, GEP has had a unique billing code

The objective of this study was to examine the agreement between administrative claims and medical record data from a large insurer on the utilization of personalized medicine technologies for women with incident breast cancer. We hypothesized that agreement may vary by factors that facilitate tracking of services in administrative claims (e.g., coding practices and the cost of the service) and by factors that enhance the completeness of medical record data (e.g., obtaining records from multiple vs. single providers of care). Using the examples of HER2 testing for trastuzumab therapy and GEP for adjuvant chemotherapy, we examined both testing and subsequent treatment because they are provided jointly as targeted therapies.

Methods

Study population

The study population included 775 women, age 35 – 65 years, with incident breast cancer diagnosed from July 1, 2006 through June 30, 2007 and three years of continuous health insurance coverage from Aetna Inc., a national health benefits company. We identified 2,121 women with incident breast cancer using established algorithms from claims data.^{4, 17} The study sample was limited to women with invasive breast cancer (n=787). We further excluded those with Stage IV cancer (n=7) and missing stage information (n=5). All women included in the analysis were confirmed as having incident breast cancer in medical records. Details about sample identification and sample characteristics can be found elsewhere.¹⁸

All study participants had health insurance coverage for HER2 tests, GEP, trastuzumab, and adjuvant chemotherapy. Aetna's coverage for GEP is limited to the Oncotype Dx Breast Cancer Assay (Redwood City, CA) technology and for individuals whose tumor meets the clinical criteria: estrogen-receptor- positive; lymph node-negative; and < 1 cm in size if HER2-positive or any size if HER2-negative, intermediate, or unknown; these results are used to guide treatment decisions.¹⁹ Women who met these clinical criteria (n=393) (hereafter referred to as the GEP sample) were examined for GEP and adjuvant chemotherapy analysis.

Data sources

We obtained data from administrative claims, plan enrollment information, and at least one medical record for all study participants.

Administrative claims—Administrative claims data are derived from claims submitted by health care providers to obtain payment for services rendered. Three components were included in our study.

- **Medical claims.** We used the medical claims as the primary data component to capture billing codes for HER2, GEP, trastuzumab, and adjuvant chemotherapy. Fully adjudicated medical claims were available from the Aetna central repository. Each claim has up to four diagnoses recorded with the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) diagnosis codes, up to four procedures recorded with ICD-9-CM procedure codes, one primary procedure recorded with Current Procedural Terminology (CPT) or Healthcare Common Procedure Coding System (HCPCS) code, revenue codes, and other relevant information such as date of service, site of service and provider specialty codes.
- **Pharmacy claims.** We also reviewed fully adjudicated pharmacy claims. Pharmacy claims data include national drug code (NDC), date the prescription was filled, days of supply, and quantity.
- **Patient enrollment file.** We also obtained data from the health plan enrollment file, which provides information on patient age, health plan type, and geographic location.

Medical Record—Requests for medical records were directed to the primary medical oncologist and the primary surgeon identified in claims, defined as the medical oncologist and surgeon providing the majority of visits to the patient during the 6 months following breast cancer diagnosis. Data were abstracted from available record(s). Among the 775 women with a medical record review, 7.6% had a record from the primary medical oncologist only, 9.7% had a record from the primary surgeon only, and 82.7% had records from both providers. Medical record data were reviewed for the clinical information up to 6 months following the breast cancer diagnosis.

We developed a standard medical record abstraction tool to collect detailed clinical information including record of HER2, GEP test, trastuzumab, or adjuvant chemotherapy. Trained data abstractors reviewed records, and we conducted quality assurance to ensure the accuracy of medical record data. Inter-rater agreement on use of testing and cancer specific variables between the abstractor and Ms. Keohane (co-author, research nurse) was 91% when a sample of 35 medical records was tested over the course of the study.

Variables

We created two sets of variables to document the use of HER2 testing, GEP, trastuzumab, and adjuvant chemotherapy according to the evidence of use in each of the data sources.

Evidence of HER2 testing, GEP, trastuzumab, and adjuvant chemotherapy in administrative claims—First, we identified the use of HER2 testing, GEP, trastuzumab, and adjuvant chemotherapy in administrative claims using relevant billing codes (Table 1). We identified HER2 testing in claims if women with incident breast cancer had a claim with CPT codes for FISH or IHC within 6 months following breast cancer diagnosis. GEP testing was identified using the specific HCPCS code. Trastuzumab therapy was identified using

either the HCPCS code in the medical claims or NDC code in the pharmacy claims. Adjuvant chemotherapy administration was identified using ICD-9-CM diagnostic and procedure codes, CPT codes for chemotherapy services and procedures, or HCPCS codes for selected chemotherapy agents.¹⁷ We did not use CPT modifiers and Diagnosis-Related Group (DRG) codes because they were not available in our database.

Evidence of HER2 testing, GEP, trastuzumab, and adjuvant chemotherapy in medical record—Second, we identified the use of HER2 testing, GEP, trastuzumab, and adjuvant chemotherapy according to the evidence in the medical record. Data abstractors performed a structured medical record review using specific definitions and directions in a coding manual (Table 1).

Statistical analysis

We calculated the utilization rate of each test and treatment from administrative claims and medical records respectively. We tested whether the proportion of women with evidence of use in the medical record equaled the proportion of women with evidence of use in the administrative claims. To determine agreement between administrative claims and medical records, we compared the evidence of use per woman. We considered two scenarios when assessing the agreement between the two data sources. First, we assumed that neither data source was perfect and employed three primary measures to assess the level of agreement: the percent overall agreement (agreement on positives and negatives), percent positive agreement (the ratio of positives identified in both data sources to the average value of positives from either data source), and percent negative agreement (the ratio of negatives identified in both data sources to the average value of negatives from either data source). Second, we treated medical record data as the gold standard and calculated the sensitivity (proportion of true positives) and specificity (proportion of true negatives), positive predictive values, and negative predictive values of the administrative claims data in identifying test and treatment utilization.

As a supplementary measure of agreement, we also calculated the kappa statistic. The kappa statistic is known to be sensitive to prevalence and unbalanced margin totals.^{20–24} We used this statistic to categorize the level of agreement based on Landis and Koch's classification²⁵: slight agreement (<0.2), fair agreement (0.21–0.40), moderate agreement (0.41–0.60), substantial agreement (0.61–0.80), almost perfect agreement (0.81–1.00).

STATA version 10 was used for all statistical analyses (StataCorp, College Station, TX). The study protocol was reviewed and approved by the Institutional Review Boards of the University of California, San Francisco and Partners HealthCare, Boston.

Results

Utilization rate of HER2, GEP, trastuzumab, and adjuvant chemotherapy

Utilization of HER2 testing was significantly different between the medical records and claims (Table 2). Ninety-seven percent of women had a documentation of HER2 testing in the medical records, while only 76.5 percent had a HER2 testing in claim ($p<0.001$). There was a trend that the GEP utilization rate in claims was higher than the rate in the medical records (30.8% vs. 24.9%, $p=0.07$). The difference in the utilization of trastuzumab and adjuvant chemotherapy between medical record and claims data was not statistically significant (13.4% vs. 11.9%, $p=0.36$ for trastuzumab; 38.9% vs. 43.3%, $p=0.22$ for adjuvant chemotherapy).

Patterns of agreement and disagreement

We found discrepancies between the medical records and claims for both testing and treatment (Table 3). HER2 testing and trastuzumab were more often documented in medical records. Approximately 20% of women had no claim consistent with a HER2 test, but had one documented in the medical record. Only 2% had a claim consistent with a HER2 test but did not have a test documented in the medical record. In contrast, we found that GEP and adjuvant chemotherapy were more often documented in claims. Seven percent of women eligible for GEP had a GEP claim without documentation in the medical record; over one percent had GEP documented in the record but not in claims.

Agreement between data sources: Testing

Agreement between data sources varied by test type (Table 4). The overall agreement was 75.2% for HER2. The kappa statistic was 0.014, categorizing the level of agreement as “slight.” When the medical record was treated as the gold standard, sensitivity, the proportion of women with HER2 in claims to all women with HER2 documentation in records, was 76.7%. Specificity, the proportion of women without HER2 in claims to all women without HER2 in records, was 29.2%.

GEP had greater agreement than HER2; the overall agreement of claims and records was 91.6%. The kappa statistic was 0.792, categorizing the GEP agreement as “substantial.” When the medical record was treated as the gold standard, sensitivity and specificity were 94.9% and 90.1%, respectively.

Agreement between data sources: Treatment

Overall, the agreement for the two treatments was high (Table 3). For trastuzumab, the overall agreement was 95.1%. When the medical record was treated as the gold standard, sensitivity was 76.0% and specificity was 98.1%.

For adjuvant chemotherapy, the overall agreement was 90.0%. When the medical record was treated as the gold standard, sensitivity was 92.1% and specificity was 87.9%. The kappa statistic categorized the agreement for both treatments as “substantial” (kappa=0.778 for trastuzumab; kappa=0.785 for adjuvant chemotherapy).

Agreement between data sources by the number of records reviewed

There was a trend that women with two records reviewed were more likely to have consistent documentation of use in both data sources (medical records and claims) than women with only one record reviewed (Table 5). Among those with only one record reviewed, there was a trend that women with a record from the oncologist were more likely to have consistent documentation of GEP use than women with a record from the surgeon.

Discussion

This study examined the agreement between two widely used and important data sources, administrative claims and medical record data, in identifying two pairs of targeted tests and treatments (HER2 and trastuzumab; GEP and adjuvant chemotherapy) for breast cancer patients. Our study contributes to the existing literature by examining personalized medicine technologies and by examining whether the agreement varies by the number of records reviewed and provider specialty. Overall we found good agreement between claims and medical records for GEP, trastuzumab, and chemotherapy, but poor agreement for HER2. A higher level of overall agreement, negative agreement, and specificity was associated with availability of unique billing codes (GEP vs. HER2, trastuzumab vs. chemotherapy) and cost

(GEP vs. HER2). There was a trend that greater agreement was associated with multiple records reviewed.

Our study observed 90%–95% overall agreement for GEP, trastuzumab, and adjuvant chemotherapy. This level of agreement in a commercially insured population is comparable to studies of the Medicare population. These studies found a high agreement between Medicare claims and medical records of hospitals or treating physicians for chemotherapy^{1, 7, 26} (94%–97% overall agreement), for cancer surgery procedures^{3, 27} (90% overall agreement, 0.70–0.90 kappa statistic), and for radiation treatment⁸ (88%–94% overall agreement).

We observed that the documentation of GEP and adjuvant chemotherapy was higher in claims than in medical records, and that about 7% of the GEP-eligible sample had a claim for GEP or chemotherapy but no documentation in the records. Similar to our observation, Du et al. found that many patients had a claim of chemotherapy but no documentation in their medical records and suggested that potential reasons may include erroneous claims or incomplete medical record data.¹ We expected that our study period would capture the majority of testing and treatment given that GEP tests were typically administered at the time of breast cancer diagnosis or shortly thereafter and that SEER-Medicare studies found most patients had a claim of cancer treatment within 2 months²⁷. However, our medical record data may have been incomplete if GEP tests were ordered by other providers, if medical records were missing laboratory reports, or if chemotherapy was administered by the medical oncologist but the record was requested from the primary surgeon only.

We found that patterns of agreement (or disagreement) varied by the type and characteristics of services. One interesting observation of the two test/treatment pairs was that higher overall agreement, higher negative agreement, and higher specificity were associated with unique billing codes (GEP vs. HER2, trastuzumab vs. chemotherapy) and costs (GEP vs. HER2). Data quality and completeness may have been associated with specific characteristics of services - costs, coverage policy, and coding. We observed lower utilization, lower overall agreement, lower negative agreement, and lower specificity in claims documentation for HER2 than for GEP. A substantial proportion (22.6% of the study sample) had either the IHC or FISH test documented in the record but not in claims. HER2 might have been less identifiable in claims than GEP because it was less expensive, lacked specific codes to distinguish IHC and FISH testing for other purposes, or represented “bundling” of codes under a larger pathology payment category. It was interesting to discover that GEP had a higher rate in claims than in the medical record. For expensive tests and treatment, costs may have created a strong incentive for developing specific codes for billing and reimbursement, thus improving the quality of administrative claims for identifying the use of such services.

Our findings supported the value of reporting overall and individual agreement measures in addition to kappa for a full evaluation of agreement and to avoid potential bias due to prevalence and sampling as recommended by other studies.^{20–24} We observed an extremely low kappa value (kappa=0.014) for HER2, a test for which 96.9% of the study sample had documentation in their medical records. Kappa may not be an appropriate measure for our study given that HER2 is recommended for all invasive breast cancer patients, and thus we expected a high prevalence rate. This observation is consistent with prior studies that demonstrate a paradoxically low kappa coexistent with high agreement, due to unbalanced margin totals (e.g., the “Yes” group in the claims and the “Yes” group in the medical records were substantially greater than 50% of the total sample).²¹

Our study had several limitations. The medical record data could have been incomplete because we did not conduct a record review of all providers. We requested data from the primary medical oncologist and the primary surgeon because their records should have captured the majority of tests and treatments for the patients and were the best choices given limited resources. However, multiple records were not available for all patients. Second, our analysis was based on fully adjudicated claims in only the Aetna system. We would not have captured the claims paid in full by another plan for women with secondary medical coverage. Our claims may have been incomplete due to coordination of benefits and the order of payment, although we expect this was uncommon. Third, we could not have distinguished whether the medical records were electronic or paper-based. Electronic medical records were expected to improve efficiency and reduce transcription errors, thereby improving the accuracy and completeness of medical record data. To compensate, we conducted training for abstractors and provided specific instructions to identify the information for abstraction. Although the time and effort to abstract the medical records may have varied between paper-based and electronic medical records, we expected minimal impact of different medical record systems on data quality and completeness for our study. Lastly, perhaps the true gold standard does not exist, given the potential incompleteness of medical records. One possible strategy to counter this potential limitation would have been to implement a standardized prospective data collection method. Such prospective data collection would have given an indication of the accuracy of both medical records and claims data.

In sum, we examined the agreement for two pairs of targeted tests and treatments between administrative claims and medical record data. We found good agreement between the two data sources for GEP, trastuzumab, and adjuvant chemotherapy, but relatively poor agreement for HER2.

For health services and outcomes research, choosing one source individually versus multiple sources will depend on the study goal, data need and availability, and the characteristics of services under consideration. Our findings demonstrate several implications on the choice of data sources for examining personalized medicine technologies. Administrative claims appear to be sufficient for tracking the use of genomic tests and treatments that have specific billing codes. Medical records are the preferred data source for studies that require either detailed clinical information unavailable in claims, or studies that examine the use of genomic tests and treatments without specific billing codes.

Our findings serve as the first step toward evaluating data sources and building an evidence base for examining the utilization of personalized medicine, as well as more general tests and treatments with similar characteristics. Emerging personalized medicine technologies may present new challenges for researchers, compared to traditional treatments and therapies, due to their evolving reimbursement and coverage policies and coding issues.²⁸ More research is needed to further examine the development of coding for emerging technologies; how such effort can contribute to building an evidence base for clinical practice and policies, and the potential impact of misclassification based on one imperfect data source on utilization and health outcomes.

Acknowledgments

Funding support: the Aetna Foundation and the National Cancer Institute (P01CA130818-01A1).

The authors thank Michele Toscano of Aetna for managing data preparation activities at Aetna, and Carolyn Jevit of Aetna for developing the medical record database in ACCESS.

References

1. Du XL, Key CR, Dickie L, Darling R, Geraci JM, Zhang D. External validation of medicare claims for breast cancer chemotherapy compared with medical chart reviews. *Med Care*. Feb; 2006 44(2): 124–131. [PubMed: 16434911]
2. Gold HT, Do HT. Evaluation of three algorithms to identify incident breast cancer in Medicare claims data. *Health Serv Res*. Oct; 2007 42(5):2056–2069. [PubMed: 17850533]
3. Miller DC, Saigal CS, Warren JL, et al. External validation of a claims-based algorithm for classifying kidney-cancer surgeries. *BMC Health Serv Res*. 2009; 9:92. [PubMed: 19500395]
4. Nattinger AB, Laud PW, Bajorunaite R, Sparapani RA, Freeman JL. An algorithm for the use of Medicare claims data to identify women with incident breast cancer. *Health Serv Res*. Dec; 2004 39(6 Pt 1):1733–1749. [PubMed: 15533184]
5. Ramsey SD, Scoggins JF, Blough DK, McDermott CL, Reyes CM. Sensitivity of administrative claims to identify incident cases of lung cancer: a comparison of 3 health plans. *J Manag Care Pharm*. Oct; 2009 15(8):659–668. [PubMed: 19803555]
6. Setoguchi S, Solomon DH, Glynn RJ, Cook EF, Levin R, Schneeweiss S. Agreement of diagnosis and its date for hematologic malignancies and solid tumors between medicare claims and cancer registry data. *Cancer Causes Control*. Jun; 2007 18(5):561–569. [PubMed: 17447148]
7. Warren JL, Harlan LC, Fahey A, et al. Utility of the SEER-Medicare data to identify chemotherapy use. *Med Care*. Aug; 2002 40(8 Suppl):IV-55–61.
8. Virnig BA, Warren JL, Cooper GS, Klabunde CN, Schussler N, Freeman J. Studying radiation therapy using SEER-Medicare-linked data. *Med Care*. Aug; 2002 40(8 Suppl):IV-49–54.
9. Tisnado DM, Adams JL, Liu H, et al. What is the concordance between the medical record and patient self-report as data sources for ambulatory care? *Med Care*. Feb; 2006 44(2):132–140. [PubMed: 16434912]
10. Schenck AP, Klabunde CN, Warren JL, et al. Data sources for measuring colorectal endoscopy use among Medicare enrollees. *Cancer Epidemiol Biomarkers Prev*. Oct; 2007 16(10):2118–2127. [PubMed: 17932360]
11. Schenck AP, Klabunde CN, Warren JL, et al. Evaluation of claims, medical records, and self-report for measuring fecal occult blood testing among medicare enrollees in fee for service. *Cancer Epidemiol Biomarkers Prev*. Apr; 2008 17(4):799–804. [PubMed: 18381471]
12. Wolff AC, Hammond ME, Schwartz JN, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *J Clin Oncol*. 2007; 25(1):118–145. [PubMed: 17159189]
13. Carlson RW, Moench SJ, Hammond ME, et al. HER2 testing in breast cancer: NCCN Task Force report and recommendations. *J Natl Compr Canc Netw*. Jul; 2006 4(Suppl 3):S1–22. quiz S23–24.
14. De Laurentiis M, Canello G, Zinno L, et al. Targeting HER2 as a therapeutic strategy for breast cancer: a paradigmatic shift of drug development in oncology. *Ann Oncol*. May; 2005 16(Suppl 4):iv7–13. [PubMed: 15923434]
15. Genomic Health Incorporated Press Release. Genomic Health Announces Fourth Quarter and Year-End 2008 Financial Results and Business Progress, Provides 2009 Financial Guidance; Redwood City, CA. Feb 3, 2009;
16. Ferrusi IL, Marshall DA, Kulin NA, Leighl NB, Phillips KA. Looking back at 10 years of trastuzumab therapy: what is the role of HER2 testing? A systematic review of health economic analyses. *Personalized Medicine*. 2009; 6(2):193–215. [PubMed: 20668661]
17. Hassett MJ, O'Malley AJ, Pakes JR, Newhouse JP, Earle CC. Frequency and cost of chemotherapy-related serious adverse effects in a population sample of women with breast cancer. *J Natl Cancer Inst*. Aug 16; 2006 98(16):1108–1117. [PubMed: 16912263]
18. Haas, J.; Phillips, K.; Liang, S., et al. Utilization of Targeted Testing Strategies & Therapies for Breast Cancer in Clinical Practice. 2010. Manuscript under review
19. Aetna. Clinical policy bulletin number 352: Tumor Markers. Hartford, CT: Aetna; 2009.
20. Chen G, Faris P, Hemmelgarn B, Walker RL, Quan H. Measuring agreement of administrative data with chart data using prevalence unadjusted and adjusted kappa. *BMC Med Res Methodol*. 2009; 9:5. [PubMed: 19159474]

21. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol.* 1990; 43(6):543–549. [PubMed: 2348207]
22. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol.* 1990; 43(6):551–558. [PubMed: 2189948]
23. Kundel HL, Polansky M. Measurement of observer agreement. *Radiology.* Aug; 2003 228(2):303–308. [PubMed: 12819342]
24. Weiner MG, Garvin JH, Ten Have TR. Assessing the accuracy of diagnostic codes in administrative databases: the impact of the sampling frame on sensitivity and specificity. *AMIA Annu Symp Proc.* 2006:1140. [PubMed: 17238759]
25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* Mar; 1977 33(1):159–174. [PubMed: 843571]
26. Chen-Hardee S, Chrischilles EA, Voelker MD, et al. Population-based assessment of hospitalizations for neutropenia from chemotherapy in older adults with non-Hodgkin's lymphoma (United States). *Cancer Causes Control.* Jun; 2006 17(5):647–654. [PubMed: 16633911]
27. Cooper GS, Virnig B, Klabunde CN, Schussler N, Freeman J, Warren JL. Use of SEER-Medicare data for measuring cancer surgery. *Med Care.* Aug; 2002 40(8 Suppl):IV-43–48.
28. Phillips KA, Liang SY, Van Bebber S. Challenges to the translation of genomic information into clinical practice and health policy: Utilization, preferences and economic value. *Current Opinion in Molecular Therapeutics.* 2008; 10(3):260–266. [PubMed: 18535933]

Table 1

Medical records definitions and claims data codes.

	HER2 test	Gene expression profiling	Trastuzumab	Adjuvant Chemotherapy
Medical record definitions	<p>“Please review the medical record to determine whether ANY HER2 test was conducted and determine the type of test and result of the test” There are two types of HER2 test – Immunohistochemistry (IHC) and Fluorescent in situ hybridization (FISH). HER2 may also be labeled as HER2/neu or c-erb-B2.</p>	<p>“Please review the medical record to determine whether ANY GEP test (including OncotypeDX, MammaPrint, and recurrence score) was conducted and determine the type of test and result of the test”</p>	<p>“Please review the medical records for use of trastuzumab (Herceptin) during the selected date range.”</p>	<p>“Please review the medical records for use of adjuvant chemotherapy during the selected date range.” Doxorubicin/cyclophosphamide (“AC”) Cyclophosphamide/methotrexate/fluorouracil (“CMF”) Fluorouracil/doxorubicin/cyclophosphamide (“CAF”) Epirubicin/cyclophosphamide (“EC”) Fluorouracil/epirubicin/cyclophosphamide (“FEC”) Doxorubicin/cyclophosphamide/paclitaxel (“AC-T”) Docetaxel/cyclophosphamide (“TC”) Docetaxel/carboplatin Docetaxel/doxorubicin/cyclophosphamide (“TAC”) Doxorubicin or epirubicin then CMF (“AorE – CMF”) Paclitaxel/albumin bound paclitaxel Vinorelbine only Capecitabine only Docetaxel only</p>
Type of Coding in Claims	IHC			
CPT	88360, 88361	NA	NA	96400, 96408, 96410, 96412, 96414, 96545
HCPCS	NA	S3854 (adopted in 2006)	J9355	J9000, J9001, J9150, J9151, J9180, J9211, J9293 (Anthracyclines) J9070, J9080, J9090 – J9097, J9280, J9290, J9291 (Alkylators) J9170, J9265 (Taxanes) J9190, J9201, J9250, J9260, J8610 (Antimetabolites) J9060, J9062 (Platinum) J9360, J9370, J9375, J9380, J9390 (Vineca alkaloids) J9010, J9015, J9020, J9031, J9040, J9045, J9050, J9065, J9100, J9110, J9120, J9130, J9140, J9165, J9181, J9182, J9185, J9200, J9201, J9202, J9206, J9208, J9209, J9211 – J9218, J9230, J9245, J9250, J9260, J9265, J9266, J9268, J9270, J9280, J9290, J9291, J9295, J9310, J9320, J9340, J9350, J9355, J9357, J9600, J9999, J8999, J8510, Q0083 – Q0085 (other)
NDC	NA	NA	50242-134-68	NA
ICD9 diagnosis	NA	NA	NA	V58.1, V66.2, V67.2
ICD9 procedure	NA	NA	NA	99.25

HCPCS= Healthcare Common Procedure Coding System; NDC=National Drug Code; CPT= Current Procedural Terminology; ICD=International Classification of Diseases, 9th revision.

Table 2

Utilization rates of HER2, GEP, trastuzumab, and adjuvant chemotherapy in the medical record and administrative claims

	Medical record N (%)	Administrative claims N (%)	P-value ^a
Testing			
HER2 testing (n=775)	751 (96.9)	593 (76.5)	P<0.001
GEP (n=393)	98 (24.9)	121 (30.8)	p=0.07
Treatment			
Trastuzumab (n=775)	104 (13.4)	92 (11.9)	P=0.36
Adjuvant chemotherapy (n=393)	153 (38.9)	170 (43.3)	P=0.22

^a: Null hypothesis: the proportion of women with evidence of use in the medical record equals the proportion of women with evidence of use in the administrative claims.

Table 3

Patterns of agreement between the medical record and administrative claims

	Patterns of agreement (Administrative claims/Medical record)			
	Yes/Yes N(%)	Yes/No N(%)	No/Yes N(%)	No/No N(%)
<i>Testing</i>				
HER2 testing (n=775)	576 (74.3)	17 (2.2)	175 (22.6)	7 (0.9%)
GEP (n=393)	93 (23.9)	28 (7.1)	5 (1.3)	267 (67.9)
<i>Treatment</i>				
Trastuzumab (n=775)	79 (10.2)	13 (1.7)	25 (3.2)	658 (84.9)
Adjuvant chemotherapy (n=393)	141 (35.9)	29 (7.4)	12 (3.0)	211 (53.7)

Table 4

Measures of agreement between the medical record and administrative claims

	Measures of agreement				Medical record = Gold Standard			
	Overall Agreement (%)	Positive Agreement (%)	Negative Agreement (%)	Kappa	Sensitivity (%)	Specificity (%)	Positive Predictive Value (%)	Negative Predictive Value (%)
<i>Testing</i>								
HER2 testing (n=775)	75.2	85.7	6.8	0.014	76.7	29.2	97.1	3.8
GEP (n=393)	91.6	84.9	94.2	0.792*	94.9	90.5	76.9	98.1
<i>Treatment</i>								
Trastuzumab (n=775)	95.1	80.6	97.2	0.778*	76.0	98.1	85.7	96.3
Adjuvant chemotherapy (n=393)	90.0	87.3	91.1	0.785*	92.1	87.9	82.9	94.6

* substantial agreement

Table 5

Agreement between the medical record and administrative claims by the number of records reviewed

	Overall agreement (%)			
	HER2 testing (n=775)	GEP (n=393)	Trastuzumab (n=775)	Adjuvant chemotherapy (n=393)
<i>One record vs. two records</i>				
One record	70.2	90.9	94.0	83.3 ²
Two records	76.3	91.7	95.3	90.8
<i>Provider specialty among one record</i>				
Oncologist	69.5	97.1 ¹	96.6	82.4
Surgeon	70.7	84.4	92.0	84.4

¹
p=0.073;²
p=0.069