POSTER ABSTRACTS
8th Annual HMO Research Network Conference
April 9-10, 2002 Long Beach, CA

Cancer
39

The Reliability of Breast Cancer Case Identification Using Claims Data

Sharon J. Rolnick, Ph.D., MPH - HealthPartners Research Foundation
Gene Hart - Group Health Cooperative
Mary B. Barton, MD - Harvard Pilgrim Health Care
Lisa Herrinton, Ph.D. - Kaiser Permanente Northern California
Sharon Kaufer Flores, MS - HealthPartners Research Foundation
Kristi J. Paulsen, BA - HealthPartners Research Foundation
Gail Husson, MPH, M.Ed. - Kaiser Permanente Northern California
Emily L. Harris, Ph.D. - Kaiser Permanente Northwest
Ann M. Geiger, Ph.D. - Kaiser Permanente Southern California
Joanne Elmore, MD, MPH - Harborview Medical Center
uzanne W. Fletcher, MD, MSc - Harvard Pilgrim Health Care

Background: The "gold standard" of cancer case identification is commonly thought to be use of the National Cancer Institute's Surveillance, Epidemiology and End Results (SEER) program. To determine the degree to which the methods developed at a non-SEER HMO could identify the same breast cancer cases captured by SEER methods, we applied an algorithm based on claims data from a non-SEER site to two sites with SEER registries.

Methods: At HealthPartners, an algorithm based on computerized health plan records from ambulatory care and hospitalizations was developed to ascertain breast cancer cases using ICD-9 codes 174.x and 233.0. Group Health Cooperative (GHC) in Seattle and Kaiser Permanente Northern California (KPNC) in Oakland duplicated the HealthPartners case identification algorithm with computerized health plan records and then compared the results to information from their SEER registries. At GHC, the algorithm was applied to the database twice – first requiring one breast cancer diagnosis, then requiring two diagnoses in the time period. At KPNC, only one diagnosis was required. Discordance between ICD-9 identification and SEER identification at GHC was assessed by examination of a random sample (N=50) of computerized encounter records.

Results: Computerized GHC records identified 2573 of 2605 (99%) SEER cases, but also identified 3345 false positives, with positive predictive value of 43%. Requiring a second notation of the diagnosis lowered sensitivity slightly (to 98%) and improved positive predictive value (to 56%). Examination of encounter records of discordant cases identified by diagnosis code but not by SEER found eight instances where SEER registry may have missed breast cancer cases. The KPNC database identified 7127 of 7445 (96%) SEER cases, but also identified 12483 false positives, resulting in a PPV of 38%.

Cancer
39

Conclusions: The method using claims data developed by HealthPartners had a high sensitivity (96-99%) when compared with SEER registries at two other CRN sites, though its positive predictive value was somewhat low at 38-56% depending on whether one or two distinct diagnoses of breast cancer were required. Identifying breast cancer cases through computerized health plan records is feasible and sensitive, but requires individual record review to eliminate "false positive" cases.