

Development and validation of case-finding algorithms for recurrence of breast cancer using routinely collected administrative data

Xu, Y¹, Kong, S¹, and Quan, ML¹¹University of Calgary

Introduction

Recurrence free survival is frequently investigated in cancer outcome studies, however is not explicitly documented in cancer registry data that is widely used for research. Patterns of events after initial treatment such as oncology visits, re-operation, chemotherapy or radiation may herald recurrence.

Objectives and Approach

This study aimed to develop and validate algorithms for identifying breast cancer recurrence using large administrative data. Two cohorts with high recurrence rates were used: 1) all young (≤ 40 years) breast cancer patients (2007-2010), and 2) all neoadjuvant chemotherapy patients (2012-2014) in Alberta, Canada. Health events after primary treatment were obtained from the Alberta cancer registry, physician billing claims, and vital statistics databases. Positive recurrence status (defined as either locoregional, distant or both) was ascertained by primary chart review. The cohort was divided into a developing (60%) and validating (40%) set. Development of algorithms geared towards high sensitivity, PPV and accuracy respectively were performed using classification and regression tree (CART) models. Key variables in the models included: a new round of chemotherapy, a second mastectomy, and a new cluster of radiologist, oncologist or general surgeon visits occurring after the primary treatment. Compared with chart review data, the sensitivity, specificity, PPV, NPV and accuracy of the algorithms were calculated.

Results

Of 606 patients, 121 (20%) had recurrence after a median follow-up 4 years. The high sensitivity algorithm had 94.2% (95% CI: 90.1-98.4%) sensitivity, 92.8% (90.5-95.1%) specificity, 76.5% (70.0-88.3%) PPV, 98.5% (97.3-99.6%) NPV and 93.1% (91.0-95.1%) accuracy. The high PPV algorithm had 74.4% (66.6-82.2%) sensitivity, 97.8% (96.5-99.2%) specificity, 90.0% (84.1-95.9%) PPV, 93.6% (91.4-95.7%) NPV and 92.9% (90.9-95.0%) accuracy. The high accuracy algorithm had 88.4% (82.7-94.1%) sensitivity, 97.1% (95.6-98.6%) specificity, 88.4% (82.7-94.1%) PPV, 97.1% (95.6-

98.6%) NPV and 95.4% (93.7-97.1%) accuracy.

Conclusion/Implications

The proposed algorithms achieved favourably high validity for identifying recurrence using widely available administrative data. Further study may be needed for improving sensitivity and PPV, and validating the algorithms in larger data for widespread use.

