

were nulliparous at diagnosis. The 5-year cumulative incidence of post-diagnosis birth was 11.7% for women diagnosed during 2007–2012 (after ASCO's 2006 guidelines), compared to 11.6% during 2000–2005 (HR = 0.98; 0.89, 1.08) and varied little by cancer type. Despite advances in fertility preservation options and recognition of fertility counseling as a part of high quality cancer care, birth rates have remained stable over the last 15 years. Low implementation of fertility counseling and limited access to fertility preservation may be contributing factors.

Published online March 2, 2017

doi: 10.1158/1055-9965.EPI-17-0024

©2017 American Association for Cancer Research.

### Identifying Patient Smoking History for Cessation and Lung Cancer Screening through Mining Electronic Health Records

Omega T, Nutter EL, Sargent J, Doherty JA, Hassanpour S

Electronic health records (EHRs) contain information about tobacco use, but smoking status and history are often inadequately captured, resulting in missed opportunities for smoking cessation intervention and lung cancer screening decision-making. Informatics methods can improve ascertainment of smoking behaviors through development of a tobacco use registry from EHRs. Methods: Using structured data and free text from our local EHR, we developed two support vector machine (SVM) models to classify smoking status (never, former, current) and smoking history (never, pack-years, cigarettes per day, years smoked). We trained and tested these models on 758 clinical notes from the Epic-based EHR of the Dartmouth-Hitchcock health system; the training set had 479 notes and the test set 280. Notes were eligible if a patient was:  $\geq 21$  years old with a clinical encounter in the EHR from 1/1/15-9/1/16. We assessed the models' performance through precision (probability that retrieved element is relevant), recall (probability that relevant element is retrieved), and the F1-score (harmonic mean of precision and recall). We also tested the models on publicly available data from the National Centers for Biomedical Computing (i2b2) Results: Of the 280 test records, 22% were current smokers, 19% former, and 59% never smokers. Accuracy assessment of our models showed: precision = 68% and recall = 85% for smoking status and for smoking history; precision = 66% and recall = 94%. The F1-scores for smoking status and history were 65% and 74%, respectively. The majority of correctly classified smokers also had one or more smoking history element ascertained with our model. Of those individuals correctly classified as never smokers ( $n = 98$ ) only two were misclassified as having a smoking history. When testing our models on i2b2 data, our F1-score was 92%. Review of misclassified records indicates that deep learning refinements to our current machine learning approach will improve performance measures. Conclusion: Machine learning models applied to our Epic EHR consistently identifies smoking history. Creating a tobacco use registry from the EHR is feasible and with advanced algorithms, will help target patients for cancer control efforts, such as smoking cessation and lung cancer screening.

Published online March 2, 2017

doi: 10.1158/1055-9965.EPI-17-0032

©2017 American Association for Cancer Research.

### Perceptions of Cervical Cancer Risk among Ohio Appalachian Women

Bernardo BM, Reiter PL, Pennell ML, Padamsee TJ, Ruffin MT, Paskett ED

OBJECTIVE: To describe how Ohio Appalachian women are over- or under-estimating risk for cervical cancer by comparing an objective measure of cervical cancer risk to subjective interpretations of risk. METHODS: A cervical cancer risk index was computed for 530 female participants recruited from clinics in Ohio Appalachia (2006–2008). Risk factors for cervical cancer, including human papillomavirus infection, number of sexual partners, parity, age at first intercourse, age at first pregnancy, oral contraceptive use and cigarette smoking were used to create a cervical cancer risk index. This index was used to categorize participants into low, average or high risk for cervical cancer. Subjective risk was assessed by asking women to indicate whether they believe their risk of cervical cancer to be below, about average, or higher than other women. Objective risk categories were compared to subjective risk to determine if women were accurately perceiving risk, or over- or under-estimating risk. A multinomial logistic regression model was used to determine correlates of risk perception accuracy. The outcome was a categorical variable consisting of over-estimating risk, underestimating risk and accurate risk perception, with accurate risk perception serving as the reference level in the model. RESULTS: Out of 530 women, 33.6% ( $n = 178$ ) over-estimated risk and 24.9% ( $n = 132$ ) under-estimated risk. Women with household incomes of  $< \$25$  k were more likely to under-estimate cervical cancer risk, as opposed to being accurate in their risk perception, than women in households earning over  $\$50$  k [OR = 2.06, 95% CI (1.04–4.07)] as were women with less than a high-school education [OR = 2.52, 95% CI (1.06–5.96)] compared to women with a college education. Also, women who reported recent alcohol use were more likely to under-estimate risk [OR = 1.94, 95% CI (1.18–3.19)] compared to women with no self-reported recent alcohol use. No variables were found to be associated with risk over-estimation. CONCLUSIONS: Many women in Ohio Appalachia do not accurately perceive their cervical cancer risk. Low income and education are associated with under-estimating risk. Future studies should determine if risk misperceptions are associated with cervical cancer screening and other preventive behaviors.

Published online March 2, 2017

doi: 10.1158/1055-9965.EPI-17-0035

©2017 American Association for Cancer Research.

### Racial/Ethnic Differences in the Use and Discontinuation of Adjuvant Endocrine Therapy by Hormone-Receptor Status in Association with Mortality among Breast Cancer Patients Enrolled in Medicare Part D

Farias AJ, Du XL

Purpose: To determine whether racial/ethnic differences in the use and discontinuation of adjuvant endocrine therapy (AET) differed by hormone-receptor status and was associated with an increased risk of mortality. Methods: We used the SEER/Medicare dataset to conduct a retrospective cohort study of women diagnosed

# Cancer Epidemiology, Biomarkers & Prevention

## Identifying Patient Smoking History for Cessation and Lung Cancer Screening through Mining Electronic Health Records

T Onega, EL Nutter, J Sargent, et al.

*Cancer Epidemiol Biomarkers Prev* 2017;26:437.

**Updated version** Access the most recent version of this article at:  
<http://cebp.aacrjournals.org/content/26/3/437.1>

**E-mail alerts** [Sign up to receive free email-alerts](#) related to this article or journal.

**Reprints and Subscriptions** To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at [pubs@aacr.org](mailto:pubs@aacr.org).

**Permissions** To request permission to re-use all or part of this article, use this link <http://cebp.aacrjournals.org/content/26/3/437.1>. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.