Prediction of Outcomes – Traditional and/ or Data Mining Methods

Westra, B.L., Delaney, C.W., Konicek, D., & Keenan, G. (2008). Nursing Standards to Support the Electronic Health Record. Nursing Outlook, 56, 258-266.e1

The purpose of this study was to discover factors for predicting hospitalization of older adults using a combination of OASIS data and Omaha System interventions. When using logistic regression for predicting hospitalization, the results were not clinically meaningful; this was thought to be due to the heterogeneity of the patients. Therefore, latent class analysis was used for clustering patients into clinically meaningful groups. Latent class analysis is a multivariate technique of clustering similar types of patients based on a combination of variables. Based on testing a combination of variables predictive of hospitalization in prior studies, 4 latent classes were identified by activity of daily living scores, who provides caregiving assistance, oral medication management, and primary diagnoses. Within each of these classes, different variables were predictive of hospitalization, either as protective or risk factors.
Westra, B.L., Savik, K., Oancea, C., Choromanski, L., Holmes, J.H., & Bliss, D. (2011). Predicting Improvement in Urinary and Bowel Incontinence for Home Health Patients Using Electronic Health Record Data. Journal of Wounds, Ostomy, & Continence Nursing, 38(1), 77 – 87

Studies were conducted comparing the effectiveness of care provided by home care clinicians using the Omaha System and Outcome and Assessment Information Set (OASIS) data. The purpose of the first study with 2,072 episodes of care from 15 home care agencies was to discover which patient and support system characteristics and interventions documented by home health clinicians were associated with improvement in urinary and bowel incontinence contrasting logistic regression PROC GEN MOD and data mining approaches. Four different data mining decision trees for each outcome were created using the J48 algorithm in the Weka software (Version 3.5.1).Using logistic regression, no patient or support system characteristics were identified that associated with improvement in either urinary or bowel incontinence, only a limited number of interventions were significant. A data mining decision tree was producible only for bowel incontinence, demonstrating a combination of patient and support system factors as well as selected interventions were important in determining whether patients would improve in bowel incontinence.

Westra, B.L., Dey, S., Fang, G., Steinbach, M., Kumar, V., Savik, K., Oancea, C., & Dierich, M. (2012). Interpretable Predictive Models for Knowledge Discovery from Homecare Electronic Health Records. M. Chyu, Ed.). Advances in Electronic Health Records (pp. 427-446) Lubbock, TX: Texas Tech University.

Improvement in oral medication management was the focus of a methodological study to compare methods of developing predictive rules that are parsimonious and clinically interpretable from electronic health record (EHR) home visit data, contrasting logistic regression with three data mining classification models. Logistic regression and three classification models using Ripper, decision trees, and Support Vector Machines were applied to a case study for one outcome of improvement in oral medication management. Predictive rules for logistic regression, Ripper, and decision trees are reported and results compared using F-measures for data mining models and area under the receiver-operating characteristic curve for all models. The rules generated by the three classification models provide potentially novel insights into mining EHRs beyond those provided by standard logistic regression, and suggest steps for further study.

Monsen, K., Westra, B.L., Oancea, S.C., Yu, Fang, Kerr, M. J. (2011). Linking Home Care Interventions and Hospitalization Outcomes for Frail and Non-Elderly. Research in Nursing & Health,34,160-168

The aims of the study were to 1) Compare the ability of four intervention data management approaches to explain hospitalization outcomes for frail and non-frail elders separately and 2) identify intervention groups associated with hospitalization for frail elders and non-frail elders. Four methods of creating intervention groups were tested: groups derived deductively for the action category, theoretical, and clinical expert consensus approaches and intervention groups derived inductively for the data-driven approach. A retrospective cohort design of electronic health record data from 14 home care agencies was used. The area under the curve (AUC) for receiver operating curves (ROC) was computed to compare the intervention management approaches for aim 1. For aim 2, logistic regression was conducted using Proc GENMOD. In the case of frail elders, AUC values ranged from .544 to .627, with the highest AUC value corresponding to the data driven model. In the case of non-frail elders AUC values ranged from .526 to .603, with the highest AUC corresponding to the clinical expert consensus model. Intervention groups varied between frail and non-frail elderly as well as different approaches to intervention groups.

Monsen, K.A., Swanberg, H.L., Oancea, C., & Westra, B.L. (2012). Exploring the Value of Clinical Data Standards to Predict Hospitalization of Home Care Patients. Appl Clin Inf ; 3: 419–436 doi:10.4338/ACI-2012-05-RA-0016

The objectives of this study were to 1) develop a measure to predict risk of hospitalization among home care patients, the Hospitalization Risk Score (HRS) using knowledge, behavior, and status scores from the Omaha System, and 2) compare it with an existing severity of illness measure, the Charlson Index of Comorbidity (CIC). Risk scores were compared using the area under the curve (AUC) for receiver operating curves (ROC) for 1,643 episodes of home care from 14 agencies. The HRS for this sample ranged from 0 to 5.6, with a median of 1.25. The CIC for this sample ranged from 0 to 9 and with a median of 0. Nearly three fourths of the sample was hospitalized at an HRS of 2, and a CIC of 1. AUC values for ROC were 0.63 for HRS and 0.59 for the CIC. The ROC curves were significantly different ($t = -7.59$, $p < 0.003$).