Journal of Clinical and Translational Science

www.cambridge.org/cts

Posters 1–50, 326 are the top 50 posters

AI and Data Science

1

Developing an annotated database to facilitate machine learning in rare pediatric cancer†

Brian Englum[1], Ashley Johnston[2], Minerva Mayorga-Carlin[1], John D. Sorkin[1] and Brajesh K. Lal[1]

[1]University of Maryland, Baltimore and [2]Georgetown University

OBJECTIVES/GOALS: Rare pediatric cancer research has been hindered by limited sample size and disease-specific databases. The training and infrastructure required to create these datasets have been cost prohibitive. Automated electronic medical record (EMR) extraction and natural language processing (NLP) hold the potential to overcome these limitations. METHODS/STUDY POPULATION: In a mixed population of adults and children with thyroid cancer, we implemented an automated data extraction process for multiple data types, including structured demographic, laboratory, and billing data and unstructured free text from pathology reports. Data extraction and cleaning software transformed EMR data into a research-ready database. NLP using regular expressions extracted meaningful, structured data from free-text pathology reports. Results from the automated data extraction and NLP were compared to manual extraction, and results were quantified using accuracy, precision, recall, F1-score, and Cohen's Kappa to account for random agreement. Additional iterations using machine learning and large language models will be performed to compare accuracy and identify an optimal strategy. RESULTS/ANTICIPATED RESULTS: From 2016 to 2023, 2,899 patients with thyroid cancer were identified. In a 50-patient subset, manual extraction of pathology reports was compared to NLP-derived data. Histology type was correctly identified in 90% of cases (kappa = 0.85), and thyroid procedure identification accuracy was 89% (kappa 0.81). Focality was 98% accurate (kappa=0.96). Vascular, lymphatic, and perineural invasion, along with extrathyroidal extension and pathologic tumor stage, were 100% accurate. Pathologic nodal staging was 96% accurate (kappa=0.89). Kappa greater than 0.8 is considered almost perfect agreement. DISCUSSION/SIGNIFICANCE OF IMPACT: Our results indicate that automated EMR extraction and NLP can be used to develop a thyroid cancer dataset that is highly detailed, accurate, and efficient to collect. After optimizing our system of data extraction, we will replicate this process in a multi-center study with the eventual goal to expand these algorithms across rare pediatric cancers.

2

Risk factors for suicide-related revisits at US Pediatric Emergency Departments†

Anna Cushing[1], Pradip P. Chaudhari[2], Ramon Durazo-Arvizu[2], Juliet Edgcomb[3], Bonnie T. Zima[3] and Amy E. West[2]

[1]Children's Hospital Los Angeles/University of Southern California; [2]CHLA/USC and [3]UCLA

OBJECTIVES/GOALS: Nearly 1 in 5 children with a suicide-related emergency department (ED) visit will have a suicide attempt within 18 months. Our objective was to determine demographic and clinical characteristics associated with time to ED revisit with self-injurious thoughts and behaviors (SITB) within 6 months. METHODS/STUDY POPULATION: We performed a retrospective cohort study of children aged 6 to 17 years with a primary mental health ED visit at 38 US children's hospitals from January 2021 to December 2023. Data were obtained from the Pediatric Health Information System administrative database. We included the first index mental health ED visit for each child. Primary outcome was time to first ED revisit with a SITB diagnosis within 6 months of a patient's index mental health ED visit, defined by ICD-10-CM diagnosis codes. We evaluated demographic and clinical characteristics associated with time to SITB revisit using a multivariable Cox Proportional Hazards Regression model. RESULTS/ANTICIPATED RESULTS: We included 201,214 unique children, 8.7% had an ED revisit with SITB within 6 months. Among children with an SITB revisit, 74.6% had SITB diagnosed at the index visit. Multiple demographic characteristics were associated with longer time to revisit including male sex (HR 0.76, 95% CI 0.74-0.79) compared to female, private insurance (HR 0.83, 95% CI 0.80-0.86) compared to public, and Hispanic race and ethnicity (HR 0.81, 95% CI 0.77-0.85) compared to non-Hispanic White. Clinical characteristics associated with a shorter time to SITB revisit included index visit diagnoses for intentional self-harm (HR 1.32, 95% CI 1.27-1.38), suicidal ideation (Hazard Ratio (HR) 1.65, 95% CI 1.59-1.71), and mood disorder (HR 1.38, 95% CI 1.33-1.44). DISCUSSION/SIGNIFICANCE OF IMPACT: Three in four children with a SITB ED revisit had SITB diagnosed at the index visit. Children with index visit suicidal ideation were at highest risk of SITB revisit. These findings highlight missed opportunities to identify and address pediatric SITB at the index ED visit and to prevent ED revisits.

3

Machine learning approaches to predict colorectal cancer screening uptake in real-world EHR data

Maliha Mehnaz Mitu[1], Tyler Thompson[2], Emily Slade[3], Charbel Salem[4], Jennifer Sword[4] and Katherine Thompson[2]

[1]University of Kentucky; [2]Dr. Bing Zhang Department of Statistics, University of Kentucky; [3]Department of Biostatistics, University of Kentucky and [4]UK King's Daughters Medical Center

OBJECTIVES/GOALS: The study's objective was to use real-world data to compare machine learning approaches for predicting colorectal cancer (CRC) screening status: (1) a Random Forest model optimized for overall accuracy and (2) Feasible Solution Algorithm (FSA) models focused on interpretability and identifying subgroups at risk of being unscreened. METHODS/STUDY

*Blue Ribbon Awardee; †Gold Ribbon Awardee

© The Author(s), 2026. The Association for Clinical and Translational Science. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 10:29:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/cts.2026.10244