ScienceDirect

Contents lists available at **sciencedirect.com**
Journal homepage: **www.elsevier.com/locate/jval**

ABSTRACT

Advanced Analytics Studies

AI1
A MACHINE LEARNING APPROACH TO PREDICTING MORTALITY IN CYSTIC FIBROSIS

Rodriguez PJ, Heagerty PJ, Goss CH, Veenstra DL, Bansal A
University of Washington, Seattle, WA, USA

Objectives: Cystic Fibrosis (CF) is a progressive genetic disease affecting the lungs. Referral for lung transplant (LTx) is recommended when patients face a high likelihood of short-term mortality. Current models for predicting mortality in CF have poor performance, making LTx referral decision-making difficult and inconsistent in practice. Our objective is to apply machine learning (ML) methods to develop a high-performing risk prediction model for short-term mortality in CF. ***Methods:*** We used data from the Cystic Fibrosis Foundation Patient Registry to develop a mortality risk prediction model for adults with CF. We used the lasso method to develop a preliminary ML model on a limited initial set of clinical and demographic variables. We evaluated the time-varying performance of baseline predictions using the area under the receiver operating curve (AUC) over time and summarized 2-year performance using the survival concordance index (c-index). Performance was compared to the existing model, forced expiratory volume in 1 second (FEV_1) alone. ***Results:*** Our lasso model identified 24 predictors of mortality from the initial limited dataset and had a higher AUC at all time points compared to the existing FEV_1 only model (c-index 0.89(0.86 - 0.92) vs 0.85(0.81 - 0.88)). ***Conclusions:*** The lasso predicted mortality better than FEV_1 alone for adults with CF in the US using a preliminary dataset. We are now training models on an expanded dataset using additional ML approaches, including ridge, elastic net, support vector machines, random forests, and boosting. Instead of choosing only one ML model, we will create an optimally weighted combination of these different models using ensemble learning methods. We hypothesize that even greater gains in performance will be achieved.

AI2
PREDICTORS OF PARKINSON DISEASE IN A MEDICARE POPULATION: AN APPLICATION OF MACHINE LEARNING IN EARLY DISEASE DETECTION

Icten Z, Burns SM, Menzin JA
Boston Health Economics, Boston, MA, USA

Objectives: Detecting Parkinson disease (PD) early in its prodromal period can facilitate timely treatment and mitigate symptoms and risks. This study aims to identify predictors of incident PD patients using Medicare administrative claims data and machine learning techniques. ***Methods:*** Medicare Part A/B claims (5% sample) from 2010Q1-2015Q4 were used to identify incident PD cases in 2015 based on ICD-9/10 diagnosis codes in patients ≥65 years old with continuous enrollment during the two-year baseline period prior to their PD diagnosis date (index date). Controls were identified in a 3:1 ratio to cases and met the same eligibility criteria, had no evidence of PD diagnosis and had a randomly selected index encounter date with the month/year matching to the case's index date. Features included demographics, comorbidities, medication and procedure utilization, and service location variables were extracted from the baseline. Data were partitioned using a 60%/20%/20% split to train, tune models and test performance on unseen data. Traditional and regularized logistic regression, k-nearest neighbor, XGBoost, support vector machine, and random forest models were built, and the best model was selected using the area under the ROC curve (AUC). Accuracy, recall, precision and F1 score were also assessed. ***Results:*** The study population included 4,575 cases and 13,725 controls (mean age=74.8 years; females=56%). The XGBoost model was the best performing model (on unseen data: AUC: 83.1%; accuracy: 79.6%; recall: 65.1%; precision: 58.2% and F1: 0.61). Features in the XGBoost model with the highest weights were age, gender, Charlson Comorbidity Index, motor symptoms (tremor, abnormal gait, involuntary movements), diagnoses related to autonomic dysfunction (urinary incontinence), diagnostic procedures (CT scan, MRI), hypertension, cancer diagnostics (mammography), ophthalmologic disease, anxiety and psychosis. ***Conclusions:*** Our study identified predictors of PD with high predictive accuracy and our findings are overall consistent with established risk factors, while indicating opportunities for further research.

AI3
PREDICTION OF BREAST CANCER USING K-NEAREST NEIGHBOUR: A SUPERVISED MACHINE LEARNING ALGORITHM

Pandey S,[1] Sharma A,[1] Siddiqui MK,[1] Singla D,[2] Vanderpuye-Orgle J[3]
[1]*Parexel International, Mohali, PB, India,* [2]*Parexel International, Bangalore, India,* [3]*PAREXEL, Glendale, CA, USA*

Objectives: Mammograms are not 100% accurate in the identifying the breast cancer. Better methods are needed to predict the breast cancer without the need of surgical biopsies. The study evaluated the prediction accuracy of breast cancer using the K-nearest neighbor (k-NN) classifier algorithm. ***Methods:*** The breast cancer dataset (containing 569 records and 32 attributes) was obtained from University of California Irvine (UCI) machine learning repository. Applying supervised machine learning technique to patient characteristics including tumor features (radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry and fractal dimension), k-NEAREST NEIGHBOUR (k-NN) was used to detect whether the mass was malignant or benign. Data were segmented into a training dataset containing the first 469 observations to build the k-NN model and a testing dataset containing the remaining observations was used to simulate new patients. Normalization of the data points was applied to rescale the features to a standard range of values. The initial choice of k = 21, approximately square root of 469 patients in our training dataset was used. Alternative k-values (k = 1,5,11,15,21,27) were also tested to optimize the model performance. The analysis was conducted using "class" package of R (v3.6.2). ***Results:*** In 100 simulations, 98% accuracy was achieved by the k-NN algorithm – i.e., only 2 out of 100, or 2 percent of masses were incorrectly classified. Choice of k=21 seems more accurate than any other choices as it has the minimum number of incorrect identification of cancerous cells. ***Conclusions:*** Supervised machine learning algorithm was shown to be capable of tackling extremely complex tasks such as identification of cancerous masses with reasonable accuracy. The application of this analysis could be an important resource for early detection of cancerous tumors and their treatment.

AI4
CAN WE DECREASE THE SCREENING BURDEN IN SYSTEMATIC REVIEWS? PERFORMANCE OF TWO NATURAL LANGUAGE PROCESSORS TO EXCLUDE RECORDS

Tetzlaff J,[1] Murad MH,[2] Wang Z[2]
[1]*Evidence Partners, Kanata, ON, Canada,* [2]*Mayo Clinic Evidence-Based Practice Center, Rochester, MN, USA*

Objectives: Systematic reviews (SR) require both comprehensive and efficient methodology and automation helps address these competing conditions. Natural language processors (NLP) may decrease screening time (e.g. prioritized screening); however, their benefits/risks as a second or autonomous reviewer require more investigation.This study assessed the performance of NLPs to exclude records from SRs and compared their performance to human reviewers. ***Methods:*** Using data from 8 completed SRs conducted by a US Evidence-Based Practice Center, we randomly selected 10% of references from each SR and trained two NLP classifiers (one support vector machine (SVM) and one Naïve Bayes (NB)) on inclusion/exclusion decisions. The classifiers screened the remaining references, leaving records unreviewed in the absence of consensus. Our primary outcome was NLP false negative (FN) rate (FN/screened records) compared to dual human-screened results. We also estimated single-human FN rates, workload savings and the potential impact of NLP FN on review results/conclusions. ***Results:*** Including 33,191 total screened records, the SRs spanned diverse topic domains (e.g. metabolic, neoplasms, respiratory), review types (e.g. interventional, umbrella, qualitative) and proportions of included studies. NLP FN rates ranged from 0 to 0.04%;

no missed studies in 6 reviews and 1 in each of 2 reviews (0.3 and 1.5% eligible studies). Estimated human FN rates were 0.11 - 0.94%. These NLPs achieved consensus exclusion for a median of 26% of records (range 3-57%). Descriptive assessment of NLP FNs suggested no or little impact on the reviews' results/conclusions. ***Conclusions:*** In these reviews, NB/SVM classifiers, running independently, incorrectly excluded fewer records than individual human reviewers. Even in this conservative scenario, these tools decreased workload to identify relevant literature. This adds to the growing body of evidence to help review teams considering such tools. Additional research spanning review topics and types will inform the use of this continually evolving technology.

Cost-Effectiveness Studies

CC1
RE-EVALUATION OF THE COST-EFFECTIVENESS OF CASCADE SCREENING AND TREATMENT STRATEGIES FOR ADULTS WITH FAMILIAL HYPERCHOLESTEROLEMIA IN THE UNITED STATES

Zawadzki N,[1] Hay JW,[1] Ahmed CD,[2] Myers KD,[2] Wilemon K,[2] McGowan MP,[3] Gidding SS,[2] Knowles JW[2]
[1]*University of Southern California, Los Angeles, CA, USA,* [2]*The FH Foundation, Pasadena, CA, USA,* [3]*Dartmouth-Hitchcock Medical Center, Lebanon, NH, USA*

Objectives: Familial hypercholesterolemia (FH) is a genetic disorder that leads to premature cardiovascular disease if untreated. Although it was previously thought that cascade genetic screening for FH is not cost-effective in the US, the cost of genetic testing has decreased substantially while PCSK9 inhibitors, a novel new lipid-lowering treatment, have been approved for use in FH. This study aimed to re-examine whether cascade genetic screening or cascade lipid screening is cost-effective for first-degree relatives of adults diagnosed with heterozygous FH in the US, compared to no cascade screening, and whether follow-on treatment with statins plus PCSK9 inhibitors (PCSK9i) is cost-effective, compared to high-intensity standard-of-care treatment with statins plus ezetimibe (HI-SOC). ***Methods:*** A decision tree was used to evaluate disease detection with the three screening strategies, while a lifetime Markov model was used to evaluate disease progression in one-year cycles until death. We report the incremental cost-effectiveness ratios (ICER) in terms of costs and quality-adjusted life years (QALYs) between six screening and treatment strategies, considering the US societal perspective. ***Results:*** At a willingness-to-pay threshold of $150,000/QALY, cascade genetic screening, whether paired with HI-SOC or PCSK9i, is the most cost-effective screening strategy, dominating all other alternatives assessed. Compared to cascade genetic screening with HI-SOC, cascade genetic screening with PCSK9i is cost-effective with an ICER of $56,034/QALY. Results were robust in one-way and probabilistic sensitivity analyses. Baseline cholesterol levels, treatment-adjusted cholesterol levels, and diagnostic test sensitivities had the greatest impact on ICERs. ***Conclusions:*** This study is the first to incorporate lower genetic testing costs, newly available information on diagnostic test accuracy, effects of novel treatments, disease prognosis, and increased cardiovascular risk in this population. The results support cascade genetic screening for FH with subsequent HI-SOC or PCSK9i treatment as an evidence-based intervention that dominates alternative strategies and provides good value relative to costs.

CC2
COST-EFFECTIVENESS ANALYSIS OF NEWBORN SCREENING AND TREATMENT FOR SPINAL MUSCULAR ATROPHY

Chen HF,[1] Hutton DW,[2] Lavieri MS,[1] Prosser LA[1]
[1]*University of Michigan, Ann Arbor, MI, USA,* [2]*University of Michigan School of Public Health, Ann Arbor, MI, USA*

Objectives: Spinal muscular atrophy (SMA), a rare pediatric disorder, was recommended for national screening in the U.S in 2018. However, the high cost of the only available drug treatment (nusinersen) and a potential future gene therapy (onasemnogene abeparvovec) raises the question of whether newborn screening for SMA will be economically favorable. The objective is to evaluate the cost-effectiveness of newborn screening and treatment of SMA in the U.S. ***Methods:*** We compared the costs and health outcomes of newborn screening (NBS) and clinical identification (CI) from a healthcare sector perspective. Treatments included standard care, drug (nusinersen, estimated $750,000, first year; $375,000, annually/patient), or gene therapy (onasemnogene abeparvovec, estimated $2,000,000/patient, one-time cost). Four strategies were compared: NBS/drug (future practice), NBS/gene therapy (future practice), CI/drug (current practice), CI/standard care (former practice). We developed a state-transition model to simulate the costs and health outcomes of a hypothetical cohort of 4,000,000 newborns over a lifetime. State-transition probabilities, costs, and health utilities were derived from published literature. Primary outcomes included costs, quality-adjusted life years (QALYs), and incremental cost-effectiveness ratios (ICERs). Costs and QALYs were discounted at 3%. ***Results:*** NBS strategies had higher costs and QALYs than CI strategies (NBS/drug: $843, 30.2440 QALYs; NBS/gene therapy: $454, 30.2442 QALYs; CI/drug: $178, 30.2428 QALYs; CI/standard care: $13, 30.2427 QALYs). When compared with CI/standard care, CI/drug had an ICER of $2,694,167/QALY. When compared to CI/drug, NBS/gene therapy had an ICER of $187,650/QALY. NBS/drug was dominated by NBS/gene therapy due to higher costs and slightly lower QALYs. However, when compared with CI/drug, NBS/drug had an ICER of $515,555/QALY. ***Conclusions:*** When compared to clinical identification with drug, newborn screening with drug or gene therapy yield cost-effectiveness results that are unlikely to be considered favorable. Future research should explore under what conditions receiving these new treatments would maximize patient benefit yet be considered more economically favorable.

CC4
COST-EFFECTIVENESS ANALYSIS OF THE STEPPED EXERCISE PROGRAM FOR PATIENTS WITH KNEE OSTEOARTHRITIS (STEP-KOA) TRIAL

Kaufman B,[1] Allen KD,[2] Coffman CJ,[2] Woolson S,[3] Bongiorni D,[3] Caves K,[4] Hall KS,[4] Heiderscheit B,[5] Hoenig H,[3] Huffman K,[4] Morey MC,[2] Ramasunder S,[3] Severson H,[6] Van Houtven C[2]
[1]*Duke University, Graham, NC, USA,* [2]*Durham VA, Durham, NC, USA,* [3]*Durham VA Medical Center, Durham, NC, USA,* [4]*Duke University School of Medicine, Durham, NC, USA,* [5]*University of Wisconsin, Madison, WI, USA,* [6]*Oregon Research Institute, Eugene, OR, USA*

Objectives: To evaluate the cost-effectiveness of the STepped Exercise Program for patients with Knee OsteoArthritis (STEP-KOA). ***Methods:*** The randomized control trial included 230 intervention and 115 control patients at two Veterans Affairs (VA) medical centers. A decision tree was used to simulate outcomes for cohorts of patients receiving Arthritis Education (control) or STEP-KOA (intervention), which consisted of an internet-based exercise training program (Step 1), phone counseling (Step 2), and physical therapy (Step 3) according to patient needs. Of the 60% (n=138) progressing to Step 2, 59% (n=81) progressed to Step 3. A VA-owned tablet with data was provided to 15% (n=35) of STEP-KOA participants. Micro-costing was used to derive costs from the VA perspective using invoices and VA labor rates. Quality-of-life (QOL) was measured using EQ-5D-5L US utility weights and differential change between arms estimated using site- and sex-adjusted linear mixed models. Incremental cost-effectiveness ratios (ICER) were calculated as the difference in costs divided by the difference in quality adjusted life years (QALY) between arms at 9 months. A Monte Carlo probabilistic sensitivity analysis (PSA) with 1000 simulations was used to generate a cost-effectiveness acceptability curve. ***Results:*** The trial found differential improvement in utility weights of 0.04 (95% confidence interval (CI) 0.003, 0.080; p=0.03) for STEP-KOA versus control at 9 months. In the base case, STEP-KOA resulted in an incremental gain of 0.03 QALYs and an incremental cost of $287 per patient for an ICER of $9,138. One-way sensitivity analyses found the largest sources of variation in the ICER were the impact on QOL and the proportion needing a VA-owned tablet. The PSA found a 99% probability of cost-effectiveness at $50,000 willingness-to-pay per QALY. ***Conclusions:*** STEP-KOA improves QOL by reducing KOA pain and has a high probability of cost-effectiveness. Resources needed to implement the program will decline as ownership of mobile health devices increases.

Comparative Effectiveness Studies & Methods

CE1
METHODOLOGICAL CHALLENGES WITH CONDUCTING NETWORK META-ANALYSES ASSESSING LONG-TERM COMPARATIVE EFFICACY IN PSORIASIS: A CRITIQUE OF ASSUMPTIONS UNDERPINNING RECENT INDIRECT TREATMENT COMPARISONS

Disher T,[1] **Peterson S,**[2] Eaton K,[1] Nair S,[3] Villacorta R,[2] Hutton B,[4] Cameron C,[5] Jansen JP[3]
[1]*EVERSANA, Burlington, ON, Canada,* [2]*Janssen Immunology Global Commercial Strategy Organization, Horsham, PA, USA,* [3]*Janssen R&D BE, Beerse, Belgium,* [4]*Ottawa Hospital Research Institute, Ottawa, ON, Canada,* [5]*EVERSANA, Sydney, NS, Canada*

Objectives: In network meta-analysis (NMA) of psoriasis trials, cross-over after an initial placebo-controlled period limits the connectivity of long-term evidence networks. We illustrate the challenges with conducting NMAs assessing efficacy beyond cross-over in psoriasis by critically appraising the assumptions underpinning recent long-term NMAs. ***Methods:*** We compared three recent NMA studies (Armstrong et al. 2019, Sawyer et al. 2018, and Diels et al. 2017) assessing long-term (beyond 16 to 24 weeks) comparative efficacy in psoriasis and investigated how the three studies derived comparative efficacy estimates in the absence of a connected network. While