5

Factors influencing clinicians' utilization of risk prediction models: an interview study

Mary Ann E Binuya, Annelies H Boekhout, Sabine C Linn, Ellen G Engelhardt, Marjanka K Schmidt

Netherlands Cancer Institute; Division of Molecular Pathology

Background: The use of risk prediction models in clinical decision-making can improve individualized care, but their adoption in clinical practice remains limited. We aimed to identify clinicians' criteria for utilization of prediction models.

Methods: We conducted 16 semi-structured interviews with medical oncologists, radiation oncologists, radiologists, surgical oncologists, clinical geneticists, and nurse specialists specialized in breast cancer from eight sites across the Netherlands. Thematic analysis was used to qualitatively summarize the interviews.

Results: Eight key determinants of model use were identified: accessibility, cost, understandability, acceptability, accuracy, actionability, risk communication benefit, and relevance to current practice. Clinicians primarily used models that were available as an online tool. Cost consideration was relevant when performing expensive, non-reimbursable, tests (e.g., gene signatures) was necessary alongside or as part of the risk calculation. Another common theme was understandability, driven by clear variable definitions, disease context, user interface, and output presentation. Acceptability by peers was also a recurring theme, with clinicians opting to use models that were used by their colleagues or presented in conferences. Clinicians' perception of accuracy was dependent on both scientific evidence (e.g., validation studies in specific cohorts) and subjective assessment (i.e., the concordance of risk estimates from the model and the clinician's personal risk assessment). Models were more likely to be used if they facilitated decision-making (actionability) or risk communication. While validity and clinical usefulness as constructs were broadly discussed, there was little direct mention of relevant statistical measures or their minimum requirements. Finally, clinicians preferred models that were developed or updated with recent data.

Conclusion: From the clinicians' perspective, use of prediction models follows a combination of practical and subjective considerations. Model developers should consider these factors when seeking to translate their models in clinical practice. Further research is needed to examine the magnitude of impact of each factor on model use.

Conflicts of interest to disclose: We declare no competing interests

5

Information bias in estimating vaccine effectiveness due to informed consent for COVID-19 vaccine register

C Henri van Werkhoven, Brechje de Gier, Scott McDonald, Hester E de Melker, Susan JM Hahné, Susan van den Hof, Mirjam J Knol

RIVM (National Institute for Public Health and the Environment);University Medical Center Utrecht; Center for Infectious Disease Control / Julius Center for Health Sciences and Primary Care

Introduction: National vaccine registers are essential for high quality monitoring and evaluation of vaccination programs. Requirement of informed consent for registration of vaccinations causes non-consenting individuals to be classified as unvaccinated, leading to information bias. We quantified the bias of vaccine effectiveness (VE) and provide bias-corrected estimates.

Methods: We used national COVID-19 vaccination and hospitalization data from the period dominated by the Delta variant before the booster campaign (11 July to 15 November 2021). VE against COVID-19 hospitalization and ICU admission, calculated as (1-relative risk)*100%, was estimated for individuals 12-49, 50-69, and ≥70 years of age using negative binomial regression corrected for age and date. Informed consent rates by covariates birth year, sex, region, and date were based on anonymous data of individuals vaccinated by the Municipal Health Services (GGD) (86% of vaccinations) and extrapolated to vaccinations by other providers. To estimate corrected VEs, we iteratively assigned hospitalized patients without vaccination record a probability of being vaccinated, based on covariate dependent vaccination uptake, consent rate and VE estimated in the previous iteration, until convergence occurred. Absolute bias was calculated as change in VE; relative bias as uncorrected/corrected relative risk.

Results: In total 8,804 hospitalizations and 1,692 ICU admissions were observed. Both absolute and relative bias due to non-consent was largest in the 70+ age group. In this group, non-consent was 7.0%; VE of full vaccination with primary series against hospitalization was 75.5% (95%CI 73.5; 77.4) before and 85.9% (95%CI 84.7; 87.1) after correction (absolute bias -10.4 percentage point, relative bias 1.74); VE against ICU admission was 88.7% (95%CI 86.2; 90.8) before and 93.7% (95%CI 92.2; 94.9) after correction (absolute bias -5.0%, relative bias 1.79).

Conclusion: Modest non-consent in registration data results in substantial underestimation of VE estimates. Known non-consent rates can be used to estimate and correct for this bias.

Conflicts of interest to disclose: We declare no competing interests

5

Real-World Data as supplementary controls for the prospective randomized HOVON-103 trial in intensively treated elderly acute myeloid leukemia patients

Sjoerd J.F. Hermans, Yvette van Norden, Erik D. van Werkhoven, Avinash G. Dinmohamed, Peter C. Huijgens, Gert J. Ossenkoppele, Gerwin A. Huls, Bob Löwenberg, Thomas Pabst, Dimitri A. Breems, Douwe Postmus, Francesco Pignatti, Jurjen Versluis, Jan J. Corne

Erasmus Medical Center Rotterdam; Department of Hematology, Cancer Institute

Introduction: To replace or supplement control patients in prospective phase II/III studies, it has been suggested to use Real-World Data (RWD). However, it is unknown how outcome of RWD-patients compares to outcome of trial control patients. Here, we compared the outcome of control patients of the prospectively randomized phase II HOVON-103 (H103) trial for elderly, intensively treated AML-patients with the outcome of similarly treated elderly AML-patients registered by the Netherlands Cancer Registry (NCR).

Methods: The H103 randomized 679 patients between standard first-line intensive chemotherapy with or without selinexor, tosedostat, or lenalidomide, in 3 consecutive randomizations. We retrospectively identified 2,469 newly diagnosed elderly AML-patients between 2014-2018 from the NCR, of whom 322 were selected for comparison with all H103-controls (n=279) based on first-line intensive treatment, WHO performance score ≤2 or unknown WHO status, and time between diagnosis and treatment ≤33 days. We performed 1:1 nearest neighbor propensity score matching (PSM) for age, modified European LeukemiaNet (ELN) 2017 AML risk classification, and white blood cell count.

Results: Before PSM, NCR-patients compared well with H103-controls for age (median: 69 vs. 69 years, P=0.51), sex (61.5% vs. 61.6% male, P=1.00), blood parameters, and modified ELN2017 risk classification (16.8% vs. 13.3% favorable, P=0.43). WHO performance status was missing in 53.7% of NCR-patients vs. 1.4% in H103-controls. Comorbidity and toxicity data were absent in the NCR. Patient characteristics were also balanced after matching. After PSM, 2-year OS was 29±3% for NCR-patients vs. 39±3% for H103-controls.

Conclusion: Selected NCR-patients were comparable with H103-controls for most patient characteristics present in both databases. However, a significantly lower OS was observed for NCR-patients. The lack of recorded performance status and absence of comorbidity and toxicity scores in NCR-patients might explain these differences. Unless RWD AML are supplemented with these important scores, RWD cannot fully replace phase II/III trial control patients.

Conflicts of interest to disclose: We declare no competing interests

5

Comparing uncertainty in individual probability predictions with various models and model average: A case study

Haoyue Wang, Junfeng Wang, Veerle Coupé

Amsterdam University Medical Center- location VUmc; Department of Epidemiology and Biostatistics

Introduction Clinical prediction models (CPMs) are frequently developed to predict risks of outcomes for patients, while the predicted outcome probability will vary depending on different sample sizes or different model development methods. We refer to this problem as model uncertainty and assess it by the standard deviation of predicted outcome probabilities of CPMs. The objectives of this study: 1) to establish whether increasing sample size of a dataset used for model development reduces prediction uncertainty; 2) to explore the relation between model types and prediction uncertainty; 3) to establish whether prediction uncertainty can be decreased by averaging over different models.

Methods Random samples with sample sizes of 500, 1000, and 5000 were drawn from a clinical dataset. Each dataset was bootstrapped 100-fold, with each different model type developed on each of the 100 bootstrapped datasets and subsequently applied to the original dataset to obtain individual risk predictions. The models under investigation included logistic regression (LR), random forest (RF), support vector machine (SVM), XGBoost and the average of all models. We calculated the standard deviation of the predicted probabilities from 100 bootstrap models for each individual and take the average of standard deviations of all individuals as the measure of prediction uncertainty.

Results At a sample size of 500, mean standard deviations for LR, RF, SVM, XGBoost and the average of these models were 0.0234, 0.0417, 0.0050, 0.0384and 0.0223, respectively; at sample size 1000, were 0.0151, 0.0368, 0.0028, 0.0347 and 0.0185, respectively; at sample size of 5000, were 0.0076, 0.0430, 0.0012, 0.0355and 0.0182, respectively.

Conclusion With increasing samples size, the prediction uncertainty decreases for LR, SVM and the average of all models, but not for RF nor XGBoost. We did not find evidence that prediction uncertainty is related to model type. The model average also provided average level of prediction uncertainty.

Conflicts of interest to disclose: We declare no competing interests

5

Using an epidemiological lens to look at the reproducibility crisis

Michiel R de Boer

University Medical Center Groningen; Department of Primary- and Long-Term Care

Introduction: Reproducibility and replicability are central tenets of scholarly research. However large scale replication projects have reported replication rates in psychology and (bio)medicine of somewhere between 40-60% depending on the exact definitions used. These results pose us to wonder whether we should embrace the uncertainty inherent in any finding from single studies, fight the replication crisis by stimulating open and transparent workflows, explore factors that could contribute to the low replication rates reported, or do all of the above. In this presentation I will focus on exploring factors that could contribute to low replication rates, revisiting the core concepts of epidemiology: validity, precision and generalizability.

Methods: A non-systematic review of the literature was conducted with the aim to retrieve replication studies and studies in which explanatory factors of low replication rates were examined. Explanatory factors that were reported were translated into epidemiological jargon where necessary.

Results: Most of the literature seems directed at the impact of the way studies are selected for replication projects (usually restricted to statistically significant results) in combination with the precision of these studies and their replications. It shows that replication rates in these projects can usually not be expected to be much higher than 60%. Most of the other studies have focused on generalizability of study findings, with conflicting results on the extent to which it plays a role in replication (failure). Studies on aspects of validity as explanatory factors for low replication rates are scarce.

Conclusion: The core epidemiological concepts validity, precision and generalizability provide a useful lens for explaining uncertainty related to non-replicated, but also replicated findings. Future studies should primarily be directed at studying validity and generalizability in relation to replication rates. The newly developed Dutch Reproducibility Network could facilitate epidemiologists and other scholars in conducting these studies.

Conflicts of interest to disclose: I am steering group lead for the Dutch Reproducibility Network (under development)

5

Estimating diagnostic accuracy under uncertainty about target disease status: a sepsis case study

Kevin Jenniskens, Christiana A Naaktgeboren, Maarten van Smeden, Lotty Hooft, Karel GM Moons, Jan Jelrik Oosterheert, Karin HAH Kaasjager, Jan-Willem van Uffen, Nandini Dendukuri, Johannes B Reitsma

University Medical Center Utrecht; Department of Epidemiology & Health Economics

Introduction: Expert panels, frequently used as reference standard in diagnostic research, typically classify the target condition dichotomously (i.e. present or absent) for each study participant. This may however lead to biased estimates of diagnostic accuracy when experts are uncertain about this classification. Eliciting probabilistic estimates may provide a solution. We provide a method on how probabilistic estimates of target condition presence, elicited from an expert panel, can be used to calculate diagnostic accuracy.

Methods: The SPACE study, investigating value of various sepsis prediction models (SIRS, qSOFA, MEWS, CBJ) in the emergency room, was used as case study. Both dichotomous and probabilistic estimates of sepsis presence were obtained from an expert panel. Diagnostic accuracy of prediction models was calculated through both a (traditional) dichotomous sepsis classification approach, and Bayesian approach using probabilistic estimates of sepsis presence.

Results: The analysis included 390 study participants, of which 79 (20.3%) were considered to have sepsis according to dichotomous classification. However, the panel expressed considerable uncertainty regarding sepsis presence (probabilities 0.2-0.8) in 65% of patients, indicating that sepsis diagnosis is difficult. The dichotomous approach yielded different diagnostic accuracy estimates compared to the Bayesian approach. Direction and size differed depending on the predictions model under study and diagnostic accuracy measure of interest. For example, sensitivity and specificity of SIRS were 95% and 46% using the dichotomous approach, and 99% and 60% using the Bayesian approach. Whereas sensitivity and specificity of CBJ, was 42% and 97% for the dichotomous approach, and 35% and 100% using the Bayesian approach.

Conclusions: Probabilistic estimates of target condition presence, elicited from expert panels, provide valuable insight in uncertainty that is normally ignored in dichotomous target condition classification by panels. The Bayesian approach allows for assessment of diagnostic accuracy using probabilistic estimates, yielding different estimates compared to the (traditional) dichotomous approach in this case study.

Conflicts of interest to disclose: We declare no competing interests