Simulated Application employed for evaluating models predicting prior risk using randomized controlled trial data

Guiyou Yang, Gerton Lunter, Wessel Ganzevoort, Sanne Gordijn, Henk Groen

University Medical Center Groningen; Department of Epidemiology

Introduction: For any prediction model, preliminary evaluation of clinical impact is wise before prospective impact studies. Specifically for evaluating models predicting treatment benefit, a method was proposed applying the results of randomized trials to individual patients (simulated application); for models predicting prior risk, decision curve analysis (DCA) may be useful. However, DCA has downsides, e.g., likely overestimated net benefit and non-intuitive Results. We propose to use simulated application to evaluate models predicting prior risk, and illustrate this procedure with an example.

Methods: A published model predicting maternal clinical deterioration in women with hypertensive disorders of pregnancy was used as an example. DCA showed using the prediction model led to a net benefit over "treating all/none" policies for cut-off values around 0.4. At this hypothetical cut-off value (0.4), reflecting how the harm of a false negative prediction was weighed against that of a false positive prediction, simulated application was applied to examine the impact of the model on the rate of maternal deterioration when DCA suggested benefits of using the model. Cesarean section rate was used as an alternative outcome. Fisher's exact test was used to compare the rates of outcomes between the "treating all" group and the group where treatment arm suggested by the model coincided with random allocation.

Results: Fisher's exact tests showed statistically significant differences (p value < 0.001) for both outcomes between the two groups. The rate difference (95% CI) of cesarean section was 5.0% (-0.4%, 10.5%), and that of maternal clinical deterioration was 15.6% (8.3%, 22.7%).

Conclusion: Simulated application is a feasible method to preliminarily assess the impact of models predicting prior risk. It provides easily interpretable results to clinicians. The results can be used as guidance for selective prospective impact studies.

Conflicts of interest to disclose: We declare no competing interests

Cross-sectionally vs longitudinally measured predictors for biological age

Annelot P Smit, Gerrie-Cor M Herber, Lieke M Kuiper, Bette Loef, H Susan J Picavet, WM Monique Verschuren

RIVM (National Institute for Public Health and the Environment);University Medical Center Utrecht; Centre for Nutrition, Prevention and Health Services (RIVM), Julius Center for Health Sciences and Primary Care (UMCU)

Introduction: People age differently, which led to the concept of 'biological' versus 'chronological' age. However, the added value of longitudinal over cross-sectional exposures and predictors of biological age are unknown. We studied exposome-related exposures as potential predictors of biological age, both cross-sectionally and longitudinally in men and women.

Methods: We used data from 4477 participants, aged 36-75 at round 4, of the long-running Doetinchem Cohort Study (DCS). Metabolomics were measured by Nightingale Health Ltd. Biological age was calculated with the MetaboHealth algorithm that consists of 14 metabolites. MetaboHealth was adjusted for chronological age to indicate whether respondents were biologically 'older' or 'younger' than chronological age-peers. Exposures included demographic, biological, lifestyle, and environmental factors. Cross-sectional exposures were based on exposure levels at round 4. Longitudinal exposures were based on the average exposure levels over 15 years (round 1 to 4), and the trend in the exposure over time. Random Forest was performed to identify the most important predictors for MetaboHealth in men and women. We compared prediction performance of the cross-sectional prediction model with the longitudinal prediction model via the mean square error (RMSE), explained variance (R2), and mean-absolute error (MAE).

Results: Prediction performance for MetaboHealth was similar for cross-sectional and longitudinal exposures in both men (R2 12.8 and 9.5 respectively) and women (R2 9.1 and 11.3 respectively). Biological exposures (e.g. BMI) and lifestyle factors (e.g. dietary factors) were most predictive for MetaboHealth in both men and women. Other important predictors were smoking behavior for men and contraceptive use and menopausal status for women.

Conclusion: Taking into account history of exposure levels (longitudinal) had no to little added value over cross-sectionally measured predictors in predicting MetaboHealth in the current study. The most important predictors for MetaboHealth were mainly from the biological and lifestyle domain and differed slightly between men and women.

Conflicts of interest to disclose: We declare no competing interests

Using negative control outcomes to identify unmeasured confounding in a self-controlled risk interval study evaluating the COVID-19 vaccine myocarditis safety signal

Sophie H Bots, Svetlana V Belitser, Carlos E. Durán, Judit Riera-Arnau, Anna Schultze, Davide Messina, Elena Segundo, Ian Douglas, Juan José Carreras, Patricia Garcia-Poza, Rosa Gini, Consuelo Huerta, Mar Martín-Pérez, Ivonne Martin, Olga Paoletti, Carl

Utrecht University; Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences

Introduction: The self-controlled risk interval (SCRI) is well suited for rapid safety evaluation of COVID-19 vaccines, but the lag between the pre-vaccination control window and later dose risk windows may introduce time-varying confounding. Negative control outcomes (NCOs) can assess this issue.

Methods: We used primary and secondary healthcare data from 5 European databases. Exposed cases with ≥1 year of follow-up before 1 September 2020 were eligible and data were extracted to the end of data availability (December 2021 - April 2022). Exposures were doses 1&2 of the Pfizer, Moderna, AstraZeneca, and Janssen COVID-19 vaccines. Outcomes were myocarditis and NCOs otitis externa (OE) and non-congenital valvular heart disease (VHD). Confounder profiles of these NCOs were similar to myocarditis and literature suggested no association with COVID-19 disease/vaccines.

The 60-day control window started 90 days before dose 1 and we used dose-specific 28-day risk windows. Analyses were stratified by vaccine brand, adjusted for calendar time in 30-day periods, and pooled with random effects meta-analysis. We expect a null finding for NCOs and interpret deviations from this as potential unmeasured confounding.

Results: We included 461 myocarditis, 40,732 OE, and 17,187 VHD cases. Myocarditis risk increased after Pfizer (IRR = 1.85, 95%CI 1.32-2.60) and Moderna (IRR = 2.39, 95%CI 1.40-4.10) dose 2. Vaccination did not affect OE risk with little heterogeneity between databases except for Janssen (I2 = 42%) and AstraZeneca dose 1 (I2 = 69%). VHD risk decreased after vaccination with considerable heterogeneity between databases (0% for Janssen to 93% for Pfizer dose 2).

Conclusion: VHD results suggest confounding that would underestimate myocarditis risk, but the OE findings contradict this. We hypothesise VHD could be a complication of COVID-19, making it a poor NCO. The degree of unmeasured confounding varied across databases. NCOs can be used to assess uncertainty in COVID-19 vaccine safety SCRIs.

Conflicts of interest to disclose: We declare no competing interests

When could real-world evidence inform regulatory decision making?

Marieke S Jansen, Olaf M Dekkers, Saskia le Cessie S, Helga Gardarsdottir, Lotty Hooft, Anton de Boer, Rolf HH Groenwold

Leiden University Medical Center; Department of Clinical Epidemiology

Introduction: Evidence based on randomized trials for benefit-risk assessment is still considered the cornerstone of regulatory decisions, yet 'real-world' evidence (RWE) is increasingly considered too. When, and to which extent, RWE is considered relevant by regulators likely depends on many factors. This study was aimed to identify factors that make RWE necessary or desirable to inform regulatory decision making.

Methods: A scoping review of peer-reviewed articles was conducted using several literature databases (PubMed, Embase, Emcare, Web of Science and Cochrane Library). Articles of any study design on factors related to RWE in regulatory decision making were included. After titles and abstract screenings, articles were read full-text, after which eligible articles were included in the review.

Results: The literature search yielded 646 articles, of which 91 articles were included in the review. Forty factors were identified from the literature, which were categorized into 8 overarching themes: generalizability, ethical considerations, feasibility, contextualization of investigational arm, real-world aspects of care, epidemiology of disease, health-technology assessments, and other. Feasibility (97%) and generalizability (89%) issues were most often mentioned as reasons to include RWE in regulatory decision making. Additionally, several interactions between factors were identified, e.g., diseases with high unmet need are typically also rare, creating a need for RWE from both an ethical and feasibility standpoint.

Conclusion: Many contextual factors play a role in determining the appropriateness of RWE to help inform regulatory decision making. For certain questions, a single factor on its own may not make RWE truly necessary, but instead multiple factors jointly determine the extent to which RWE is considered essential and pivotal in regulatory decision making. The current results lay a foundation for a future framework that can help researchers, the pharmaceutical industry, regulators and other stakeholders identify those scenarios when RWE can and should help inform regulatory decision making.

Conflicts of interest to disclose: This project was sponsored by the GetReal Institute.

Variation of diagnostic accuracy across different healthcare settings

Natasja D Vijfschagt, Huibert Burger, Michiel R de Boer, Marjolein Y Berger, Mariska MG Leeflang, Ann Van den Bruel, Thomas Fanshawe, Gea A Holtman

University Medical Center Groningen; Eerstelijnsgeneeskunde en Langdurige Zorg

Introduction: Tests developed in specialist care may have different accuracy when applied to primary care. Sensitivity and specificity may vary due to selection that occurs before referral to specialist care, but extensive studies evaluating this claim are lacking. We reanalyzed meta-analyses that included studies of diagnostic test accuracy in primary and specialist care to investigate variation between settings.

Methods: We systematically searched EBSCOhost MEDLINE for meta-analyses that included at least 10 original studies of the same diagnostic index test. Meta-analyses with at least three studies done in primary care and at least three studies done in specialist care were included in the re-analysis. We excluded studies without a clear setting definition. Firstly, we distinguished non-referred (primary care, community, other) and referred settings (specialist care, other), and secondly, we restricted to studies performed in countries in which primary care has a gatekeeping function. We used bivariate random-effects models in each meta-analysis, with setting as a covariate.

Results: We included nine meta-analyses evaluating 12 index tests; signs and symptoms (n=7), questionnaire (n=1), rapid tests (n=2), fecal test (n=1), and ultrasound (n=1). In 8 of 12 index tests sensitivity was higher in non-referred versus referred setting, with a range of 0.05 (abdominal pain for colorectal cancer) to 0.21 (rapid influenza diagnostic tests for Influenza infection), and in 10 of 12 index tests specificity was lower in non-referred setting, with a range of 0.01 (4 index tests) to 0.31 (paroxysmal cough in pertussis). Analyses limited to gatekeeping health care systems showed similar results.

Conclusion: In a majority of the index tests there is a difference between health care settings, in which for most tests the sensitivity was higher and specificity was lower in a non-referred setting. More research is needed into explaining the difference in diagnostic accuracy between settings.

Conflicts of interest to disclose: We declare no competing interests

External validation of recommended multimorbidity indices for predicting mortality: a head-to-head comparison

Premysl Velek, Marije J Splinter, M Kamran Ikram, Rikje Ruiter, Guy GO Brusselle, Bruno HC Stricker, Patrick JE Bindels, Maryam Kavousi, M Arfan Ikram, Annemarie I Luik, Barbara C van Munster, Evelien IT de Schepper, Lucy E Stirland, Silvan Licher

Erasmus Medical Center Rotterdam; Department of Epidemiology

Introduction: There are nearly twenty different multimorbidity indices available that aim to predict all-cause mortality. Several are widely used and were validated using independent data, yet a systematic assessment of their predictive performance is lacking. We present a framework for systematic evaluation and comparison of multimorbidity indices for predicting mortality.

Methods: We focused on seven indices as recommended by a recent (2020) systematic review of multimorbidity indices. We mapped all covariates used by the indices onto data from 14,926 participants within the population-based Rotterdam Study. We replicated the methods of each index's original development paper using the same follow-up time and a comparable age range and assessed their discriminative ability in terms of the C-statistic. We then compared these respective values against a separate "disease count" model containing the total number of co-occurring diseases (from among 10 chronic diseases) adjusted for age and sex.

Results: Here, we provide preliminary results on the external validation of two of the recommended indices (Lee, 2006; Tooth, 2008). All participants with complete data on the required covariates and comparable age for each index were included in the analyses. Therefore, sample sizes varied from 3,349 (mean age 73.5±7.7 years, 53.8% women) with 340 deaths during follow-up to 1,366 (mean age 74.7±2.8 years) women with 212 deaths during follow-up. C-statistics ranged from 0.62 (95%CI: 0.59-0.66, four-year follow-up) to 0.68 (95%CI: 0.64-0.72, six-year follow-up), with nearly identical predictive performance based on the disease count model (C-statistics: 0.61, 95%CI 0.58-0.64 and 0.67, 0.63-0.71, respectively).

Conclusion: This external validation study suggests poor performance of recommended multimorbidity indices to predict mortality, with similar performance compared to a disease count model. These findings call into question the incremental value of multimorbidity indices beyond a simple disease count index . Provided acceptance, we will present all available data, including assessment of additional indices.

Conflicts of interest to disclose: Three co-authors are members of the WEON 2023 Organising Committee (M. Arfan Ikram, Silvan Licher, Annemarie Luik)