Avoiding the pitfalls in developing and validating prediction models for rare outcomes in nested case-control

Barbara Rentroia-Pacheco[1], Domenico Bellomo[2], Inge M.M. Lakeman[3,4], Marlies Wakkee[1], Loes M. Hollestein[1,5,#], David van Klaveren[6,#],

[1]Department of Dermatology, Erasmus MC Cancer Institute, Erasmus University Medical Center, Rotterdam, The Netherlands
[2]SkylineDx B.V., Rotterdam, The Netherlands
[3]Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands
[4]Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands
[5]Department of Research, Netherlands Comprehensive Cancer Organization (IKNL), Utrecht, The Netherlands
[6]Department of Public Health, Center for Medical Decision Making, Erasmus University Medical Center, Rotterdam, The Netherlands
[#]These authors contributed equally to this work

E-mail:
b.rentroiapacheco@erasmusmc.nl, d.bellomo@skylinedx.com, i.m.m.lakeman@lumc.nl, m.wakkee@erasmusmc.nl, l.hollestein@erasmusmc.nl, d.vanklaveren@erasmusmc.nl

Category: PS01-23 Statistical and epidemiological methodology

Background: Molecular-biomarker-based prediction models, whether based on statistics or artificial intelligence, are becoming increasingly popular in the medical community as an alternative or complement to more traditional prediction tools, such as staging systems. Population-based cohorts are the preferred study design for building and validating such models, but they are generally expensive and frequently unfeasible, particularly, when the outcome of interest is rare, and when the collected data include expensive or difficult-to-obtain molecular biomarkers. As such, studies often develop or validate risk models in cohorts obtained based on data availability, rather than on the representativeness of a target population. Nested case-control (NCC) is an efficient study design for building and validating such models in a rare outcome setting, but it requires a fully enumerated source population, and appropriate methodologies to accommodate the under-sampling of the controls and any matching. Unfortunately, these methodologies are not systematically described in a form easily accessible to clinicians.

Methods: We systematically characterized how to correctly develop and evaluate the performance of prediction models in NCC cohorts, by weighing the subjects based on sampling probabilities to account for the sub-cohort sampling. We illustrated the use of weighted metrics for the NCC, with a validation of a model that predicts breast cancer development – the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA version 5) – in the population-based Rotterdam study. We used the C-index, threshold-based metrics, observed-to-expected events ratio (O/E ratio), calibration slope, and decision curve analysis as performance metrics. We compared the metrics obtained in the full cohort with those obtained in NCC cohorts sampled from the Rotterdam study, with and without a matched design.

Results: NCC cohorts of 326 women were derived from the full cohort of 4,377 women from the Rotterdam Study. Performance metrics without weight adjustment were biased on the NCC cohorts: the unweighted C-index of the BOADICEA model was 0.61 (0.58-0.63) for the unmatched design. However, with weight adjustment, the C-index in the NCC cohorts (0.65 (0.61-0.69)) corresponds to

the C-index in the full cohort (0.65 (0.62-0.69)), despite the NCC cohort being much smaller in size. Similarly, weighted adjustments of O/E ratio, threshold-based metrics, and net benefit for decision curves were unbiased estimates of the corresponding metrics in the full cohort, while the corresponding unweighted metrics were biased.

Conclusions: We provided a practical guide for clinicians on how to develop and validate prediction models in NCC cohorts. We showed that this study design is an efficient solution, in case of expensive or difficult-to-obtain biomarkers, and when the outcome is rare; but the performance metrics must be appropriately adjusted to the sampling procedure.