KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Laure Wynants, Sabine Van Huffel, Ben Van Calster, (2015), **Prediction models in multicenter studies: methodological aspects and current state of the art** Archives of Public Health, 73 (Suppl 1), O3.
Archived version	Final publisher's version / pdf
Published version	http://www.archpublichealth.com/content/pdf/2049-3258-73-S1-O3.pdf
Journal homepage	http://www.archpublichealth.com/
Author contact	Laure.wynants@esat.kuleuven.be *+ 32 (0)16 327670*
IR	url in Lirias https://lirias.kuleuven.be/handle/123456789/xxxxxx

(article begins on next page)

ARCHIVES OF PUBLIC HEALTH

ORAL PRESENTATION Open Access

Prediction models in multicenter studies: methodological aspects and current state of the art

Laure Wynants*, Sabine Van Huffel, Ben Van Calster

From Methods in Epidemiology Symposium
Leuven, Belgium. 17 September 2015

Increasingly, multicenter datasets are being used to develop or evaluate clinical risk prediction models. Such models estimate an individual's probability that a certain disease or condition is present (diagnostic model) or that an event will occur in the future (prognostic model). Although multicenter studies enhance the generalizability of the model, the clustered nature of the data poses several methodological challenges. We will provide an up to date overview of good practices to overcome these challenges.

When determining the required sample size, the number of events per candidate variable (EPV) is crucial to prevent overfitting when building a prediction model. We extend the EPV guidelines to multicenter studies, acknowledging the clustered nature of the data. During data collection, measurements of variables may differ between centers due to various reasons, such as subjectivity of measurements, differences in equipment and differences in patient populations. We show how the residual intraclass correlation can be used to quantify the intercenter variability. When building a prediction model, the clustered nature of the data should be taken into account during the data analysis, e.g. by using mixed effect models and variables at the center level. Only mixed effect regression can result in a model that is simultaneously calibrated (i.e. gives accurate predicted probabilities) at the center level and the population level. We give the example of the ADNEX model that was built to distinguish between several types of adnexal masses. In the end, the performance of models may differ between centers. We present how to evaluate the predictive performance of models in clustered data and show extensions to existing techniques to evaluate discrimination, calibration and clinical utility, among others by the use of meta-analytic techniques.

Published: 17 September 2015

doi:10.1186/2049-3258-73-S1-O3
Cite this article as: Wynants *et al.*: Prediction models in multicenter studies: methodological aspects and current state of the art. *Archives of Public Health* 2015 **73**(Suppl 1):O3.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central

KU Leuven, Leuven, Belgium

BioMed Central

© 2015 Wynants et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.