Abstract Book

International Workshop

Risk Prediction, Communication and Perception (RiPCoP)

in Health

June 13-15, 2023

Brandenburg Medical School

Neuruppin, Germany

Contents

Nicholas Wald (University College London, UK) Principles of screening.....4

Ruth Pfeiffer, Mitchell Gail (National Cancer Institute, Bethesda, US) Absolute risk: estimation, validation and applications.....5-6

Jacqueline Murphy (University of Oxford, UK) Prediction of cardiovascular disease risk based on conventional and novel risk factors in a large cohort study of Chinese adults.....7

Michael Pencina (Duke University, US) Best practices for algorithmic/AI governance for health systems.....8

Katarzyna Jozwiak (Brandenburg Medical School, Neuruppin, DE) Risk prediction of the coronary heart disease in survivors of Hodgkin Lymphoma.....9

David van Klaveren (Erasmus Medical Center, Rotterdam, NL) Predictive modeling approaches to personalized medicine: a comparison of regression-based methods.....10

Sander Roberti (Netherlands Cancer Institute, Amsterdam, NL & Brandenburg Medical School, Neuruppin, DE) Predicting breast cancer among female Hodgkin lymphoma patients treated with modern radiotherapy using radiation dose distributions from historic treatment.....11

Mitchell Gail (National Cancer Institute, Bethesda, US) Tools for contralateral prophylactic mastectomy decision making.....12

Antonis Antoniou (University of Cambridge, UK) BOADICEA: a comprehensive breast and ovarian cancer risk prediction model incorporating genetic and non-genetic risk factors.....13

Sam Finnikin (University Birmingham, UK) Applying risk scores to practice: The challenges and pitfalls.....14

Claudia Schneider (University of Cambridge, UK) Communicating empirical evidence: uncertainty, quality and trust.....15

Christin Ellermann (Harding Center for Risk Literacy, University of Potsdam, DE) Risk communication for disadvantaged groups.....16

Ewout Steyerberg (Leiden University Medical Center, NL) & Ben van Calster (KU Leuven, BE) Clinical prediction models: development and validation.....17

Donna Ankerst (Technical University Munich, DE) Globally-accessible individual-tailored risk prediction.....18

Ruth Pfeiffer (National Cancer Institute, Bethesda, US) Accommodating population differences in model validation.....19

Yuwei Wang (Netherlands Cancer Institute, Amsterdam, NL) External validation and clinical utility assessment of the PREDICT breast cancer prognostic model in young patients with node-negative breast cancer.....20-21

Sarah Booth (University of Leicester, UK) Methods to account for improvements in survival when developing prognostic models.....22

Anika Hüsing (University Essen, DE) Validation and improvement of 20 years prediction of cardiovascular events with information from cardiac imaging……23

Silke Schicktanz (University of Göttingen, DE) The ethical challenge of dementia prediction……24

Nicholas Wald (University College London, UK) How risk reduction should be expressed in primary prevention……25

Charlotte Dries (Harding Center for Risk Literacy, University of Potsdam, DE) Communication of epistemic uncertainty around evidence and trust……26

Martin Wolkewitz (University of Freiburg, DE) Predicting clinical outcomes of COVID-19 patients with routine data……27

Ewout Steyerberg (Leiden University Medical Center, NL) Uncertainty in predictions……28

Derek Yves Hazard (University of Freiburg, DE) Performance measures for the external validation of a multi-state prediction model for the clinical progression of hospitalized COVID-19 patients……29

Doranne Thomassen (Leiden University Medical Center, NL) Effective sample size: expressing individual uncertainty in predictions……30

Nora Pashayan (University College London, UK) Clinical utility of risk - stratification for cancer screening at the population level……31

Katie Kerr (University of Washington, USA) Recalibrating risk models for maximum net benefit……32

Mary Ann Binuya (Netherlands Cancer Institute, Amsterdam, NL) Factors influencing clinicians' utilization of risk prediction models: an interview study……33

Elena Sophia Doll (University of Heidelberg, DE) Family decisions in genomic newborn screening: Subproject Medical Psychology in the NEW_LIVES project……34

Anja Schneider (German Center for Neurodegenerative Diseases, Bonn, DE) Prediction of Alzheimer's disease……35

Ineke Bolt (Erasmus Medical Center, Rotterdam, NL) Ethics of prediction in neurodegenerative diseases……36

Irmak Begum On (University of Munich, DE) Prognostic prediction in relapsing-remitting multiple sclerosis – the way forward?……37

Ben van Calster (KU Leuven, BE) Peak performance: does a simple statistician understand random forests for risk prediction?……38

Cornelia Fütterer (Technical University of Munich, DE) Decreasing complexity of risk prediction models by introducing Discriminative Power Lasso……39

Carolin Malsch (University of Greifswald, DE) How to statistically model biologic interactions……40

Short Course 1 “Principles of Screening”

Nicholas Wald, University College London, UK

The course will be given in four parts with short breaks between each part:

Part 1 Screening using a single marker

Part 2 Screening using multiple markers

Part 3 Why cholesterol, blood pressure and polygenic risk scores are poor predictors of IHC

Part 4 Age as a screening test

Practical examples will be used from the field of prenatal screening and adult screening, with a focus on screening for future heart attacks and students.

Short Course 2 "Absolute risk: estimation, validation and applications"

Mitchell H Gail, M.D, Ph.D.
Senior Investigator
Biostatistics Branch
Division of Cancer Epidemiology and Genetics
National Cancer Institute, NIH, USA
Email:gailm@mail.nih.gov

Ruth Pfeiffer, Ph.D.*
Senior Investigator
Biostatistics Branch
Division of Cancer Epidemiology and Genetics
National Cancer Institute, NIH, USA
Email: pfeiffer@mail.nih.gov

*contact person

Abstract: Absolute (or "crude") risk is the probability that an individual who is free of a given disease at an initial age, a, will develop that disease in the subsequent interval $(a, t]$. Absolute risk is reduced by mortality from competing risks. Models of absolute risk that depend on covariates have been used to design intervention studies, to counsel patients regarding their risks of disease and to inform clinical decisions. This course will define absolute risk and discuss methodological issues relevant to the development and evaluation of risk prediction models. Various study designs and data for model building will be presented, including cohort, nested case-control, and case-control data combined with registry data. Issues relating to the evaluation of risk prediction models and the strengths and limitations of risk prediction models for various applications will be discussed. Standard criteria for model assessment will be presented, as well as loss function-based criteria applied to the use of risk models to screen a population and the use of risk models to decide whether to take a preventive intervention that has both beneficial and adverse effects. Methods for validating models in independent data when some predictors are missing are presented. Reproducibility and transportability of models are defined and criteria to assess them are presented. Finally, updating risk models when information on new (molecular) predictors become available will be discussed.

Course prerequisites: The course attendees should have a knowledge of basic statistics, epidemiologic designs, and some familiarity with survival analysis.

Learning objectives: The attendees of the course will:

- Learn what "absolute risk " (or "crude risk" or "cumulative incidence") is
- Learn how to estimate it from data from various designs
- Learn how to assess the validity and usefulness of a model of absolute risk
- Learn about applications of absolute risk models for medical and public health decision-making
- Learn how to update models with new (molecular) information

Based on the book: Ruth M. Pfeiffer, Mitchell H. Gail. Absolute Risk: Methods and Applications in Clinical Management and Public Health. Chapman and Hall/CRC, 2018

Biography Mitchell Gail:

Dr. Gail received an M.D. from Harvard Medical School and a Ph.D. in statistics from George Washington University. He is a Fellow and former President of the American Statistical Association and an elected member of the National Academy of Medicine of the National Academy of Sciences. He was named an NIH Distinguished Investigator in 2019. Dr. Gail is currently collaborating on studies of vaccines against human papilloma virus to prevent cervical cancer, and he develops methods to improve the design and analysis of epidemiologic studies. A long-standing interest is in models to predict the risk of breast cancer, including NCI's Breast Cancer Risk Assessment Tool.

Biography Ruth Pfeiffer

Dr. Pfeiffer is a tenured Senior Investigator in the Biostatistics Branch of the Division of Cancer Epidemiology and Genetics (DCEG), National Cancer Institute (NCI). She has built several absolute risk models and helped design web-based versions, including NCI's Colorectal Cancer Risk Assessment Tool. She has developed novel methods for building, assessing and updating risk prediction models, and co-authored a book on the topic with Mitchell Gail.

Prediction of cardiovascular disease risk based on conventional and novel risk factors in a large cohort study of Chinese adults

Jacqueline Murphy (DPhil candidate, Nuffield Department of Population Health, University of Oxford)
Robert Clarke (Professor, Nuffield Department of Population Health, University of Oxford)
Derrick Bennett (Associate Professor, Nuffield Department of Population Health, University of Oxford)
Sofia Massa (Lead Statistician, Oxford Clinical Trials Research Unit, University of Oxford)

BACKGROUND: Cardiovascular disease (CVD) is a leading cause of health burden globally, of which China is a primary contributor. The age-standardised prevalence of CVD in China increased by 14.7% between 1990 and 2016 (Global Burden of Diseases). Risk scores are widely used for primary CVD prevention, however scores developed in non-Chinese populations may not be appropriate for use in China.

OBJECTIVES: 1) Assess the predictive value of blood lipids, carotid plaque and electrocardiogram (ECG) measures for CVD in Chinese adults; 2) Develop a 10-year CVD risk score for China and evaluate its performance.

DATA: The China Kadoorie Biobank (CKB) is a prospective cohort study of 0.5 million adults recruited between 2004 and 2008 in 10 diverse regions of China (www.ckbiobank.org).

METHODS: Sex-specific Cox proportional hazards models stratified by region were used to predict CVD risk, and were evaluated using discrimination (c-index), calibration, and reclassification metrics. Models incorporating blood lipids, carotid plaque, and ECG were compared to assess predictive value in a subset of participants with available data. For the 10-year score, automated variable selection was carried out based on BIC and bootstrap internal validation was used to evaluate optimism due to overfitting. Methods for combining regional data into a single score, such as weighting the regional baseline survival, were compared.

RESULTS: The subset of participants with blood lipids, carotid plaque, and ECG data comprised 6,118 men and 10,782 women with median follow-up 4.8 years. These measures did not substantially improve risk prediction in the short term (4 years) in this cohort. For the 10-year score, the full cohort comprised 200,016 men and 289,356 women with median follow-up 11.0 years. The 10-year CVD incidence for men/women ranged from 6.7% to 26.4% / 4.6% to 22.2% by region. C-index statistics for men/women were 0.770/0.773 overall, and ranged from 0.733 to 0.803 / 0.738 to 0.815 by region. Correction for optimism had unsubstantial impact on the models. Calibration accuracy varied between regions.

CONCLUSIONS: This study demonstrates the importance of using data which reflect the substantial heterogeneity in CVD risk within China for development and/or validation of CVD risk scores in this setting.

Best practices for algorithmic/AI governance for health systems

Michael Pencina, Duke University, USA

Risk prediction algorithms offer tremendous opportunity to improve health and delivery of care. However, in the last decade we entered a wild-west of algorithms, with hyperactive development and insufficient focus on evaluation, implementation and monitoring. Health Systems are ill-equipped to identify, select and govern tools that offer the best promise.
In this talk we present principles for evaluation and governance of clinical decision support algorithms intended for implementation in health systems. We propose a people, process and technology framework that enables responsible application of health algorithms. Its functioning is illustrated using examples from implementation in the Duke University Health System.

Risk prediction of coronary heart disease in survivors of Hodgkin lymphoma

Simone de Vries,[1] Miriam L. Haaksma,[2] Katarzyna Jóźwiak,[3] Michael Schaapveld,[1] David C. Hodgson,[4] Pieternella J. Lugtenburg,[5] Augustinus D.G. Krol,[6] Eefke J. Petersen,[7] Dick Johan van Spronsen,[8] Sameera Ahmed,[4] Michael Hauptmann,[3] Berthe M.P. Aleman,[9] Flora E. van Leeuwen[1]

1. Department of Epidemiology, The Netherlands Cancer Institute, Amsterdam, The Netherlands
2. Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, The Netherlands
3. Institute of Biostatistics and Registry Research, Brandenburg Medical School Theodor Fontane, Neuruppin, Germany
4. Department of Radiation Oncology, Princess Margaret Cancer Centre, Toronto, Canada
5. Department of Hematology, Erasmus MC Cancer Institute, University Medical Center, Rotterdam, The Netherlands
6. Department of Radiation Oncology, Leiden University Medical Center, Leiden, The Netherlands
7. Department of Hematology, University Medical Center Utrecht, Utrecht, The Netherlands
8. Department of Hematology, Radboud University Medical Center Nijmegen, Nijmegen, The Netherlands
9. Department of Radiation Oncology, The Netherlands Cancer Institute, Amsterdam, The Netherlands

BACKGROUND: Previously developed models to predict cancer treatment-related cardiovascular diseases can be used for estimating individual risk among childhood cancer survivors. We aimed to develop prediction models to obtain absolute risk of coronary heart disease (CHD) for survivors of adolescent/adult Hodgkin lymphoma (HL).

METHODS: Prediction models were developed using a multicenter cohort of 1,433 5-year HL survivors who were treated at ages 18-50 in the Netherlands between 1965 and 2000, and who had complete data on radiotherapy field and prescribed doses, and cardiovascular follow-up. Using cause-specific hazard models, covariate-adjusted cumulative incidences for CHD were estimated in the presence of competing risks of death due to other causes than CHD. Age and smoking status at HL diagnosis, sex, and radiotherapy were included as predictors. The models were internally and externally validated. External validation was performed using a Canadian cohort of 708 HL survivors treated at ages 18-50 between 1988 and 2004.

RESULTS: After a median follow-up time of 24 years, 341 survivors were diagnosed with CHD. CHD risks at 20 and 30 years after treatment were predicted with moderate overall calibration (E/O: 0.89) and discrimination (AUCs: 0.73-0.74), which was confirmed by external validation (AUC: 0.74). Based on the model with prescribed mediastinal radiation dose, 30-year risks ranged from 4% to 78%, depending on risk factors.

CONCLUSION: We developed and validated prediction models for CHD with moderate overall calibration and discrimination. These models can be used to identify HL survivors who might benefit from targeted screening for CHD and early treatment for CHD risk factors.

Predictive modeling approaches to personalized medicine: a comparison of regression-based methods

David van Klaveren[1,2], Ewout W. Steyerberg[3], David M. Kent2

[1]*Department of Public Health, Erasmus MC University Medical Center, Rotterdam, The Netherlands*
[2]*Predictive Analytics and Comparative Effectiveness Center, Tufts Medical Center, Boston, USA*
[3]*Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands*

The benefits and harms of medical treatments vary substantially between individual patients. Predictive modeling approaches to personalized medicine are designed to predict the benefit of one treatment over another for individual patients. We aimed to compare different regression-based modeling approaches, through simulations and a case-study.

We simulated trial samples (n = 3,600; 80% power for a treatment odds ratio of 0.8) from a superpopulation (N = 1,000,000) with 12 binary risk predictors, both without and with six true treatment interactions. We assessed predictions of treatment benefit for four regression models: a "risk model" (with a constant effect of treatment assignment) and three "effect models" (including interactions of risk predictors with treatment assignment). The risk modeling approach was well-calibrated for treatment benefit, whereas effect models were consistently overfit, even with doubled sample sizes. Penalized regression reduced miscalibration of the effect models considerably. In terms of the benefit prediction error, the risk modeling approach was superior in the absence of true treatment effect interactions, whereas penalized regression was optimal in the presence of true treatment interactions.

The recently proposed Syntax Score II (SSII)-2020 was developed to predict the difference in 10-year mortality when treating complex coronary artery disease patients with heart bypass surgery rather than coronary stenting. Cox regression was first used in the SYNTAX trial data (n=1,800) to develop a prognostic index (PI) for mortality over a 10-year horizon consisting of 7 clinical predictors of mortality. Second, a Cox model was fitted which included the treatment, the PI and pre-specified treatment interactions with type of disease and with anatomical disease complexity. In contrast to its more flexible predecessor SSII-2013 which included 8 treatment interactions, SSII-2020 was well calibrated for treatment benefit at 10 years post-procedure, both at cross-validation in the same data and at external validation in new data.

The simulations and the case study both showed that robust modeling approaches – only including plausible treatment interactions – may lead to better predictions of treatment effect. Future research could focus on robust approaches for data-driven selection of treatment interactions.

Predicting breast cancer among female Hodgkin lymphoma patients treated with modern radiotherapy using radiation dose distributions from historic treatments

Sander Roberti, MSc (1), Flora E. van Leeuwen, PhD (1), Ibrahima Diallo, PhD (2), Florent de Vathaire, PhD (2), Wendy M. Leisenring, ScD (3), Rebecca M. Howell, PhD (4), Gregory T. Armstrong, PhD (5), Chaya S. Moskowitz, PhD (6), Nicola S. Russell, PhD (1), Ruth M. Pfeiffer, PhD* (7), Michael Hauptmann, PhD* (8)

(1) The Netherlands Cancer Institute, Amsterdam, The Netherlands; (2) INSERM U1030, Gustave Roussy, Université Paris-Saclay, Villejuif, France; (3) Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America; (4) University of Texas M.D. Anderson Cancer Center, Houston, Texas, United States of America; (5) St Jude Children's Research Hospital, Memphis, Tennessee, United States of America; (6) Memorial Sloan Kettering Cancer Center, New York, New York, United States of America, (7) National Cancer Institute, Rockville, Maryland, United States of America; (8) Brandenburg Medical School Theodor Fontane, Neuruppin, Germany

*authors contributed equally

Background: Historic chest radiotherapy (RT) strongly increases subsequent breast cancer (BC) risk among female Hodgkin lymphoma (HL) survivors. Accurate BC risk prediction is important to identify high-risk subgroups and aid treatment planning. Using radiation dose distributions may allow more accurate predictions for patients treated with modern techniques.

Methods: We modeled relative risks (RRs) for BC in a case-control sample (170 cases, 456 controls), nested in a Dutch cohort of 5-year HL survivors (treated 1965-2000). Dose to five locations in both breasts (central portion, four quadrants) was reconstructed. The linear excess relative risk (ERR) was estimated as $RR=1+\beta\ Dose$ with location-specific radiation dose. Absolute BC risk, accounting for competing risks, was estimated by combining RRs with age-specific BC incidence from the cohort (model M1), and was compared to a model that only incorporated mean dose to the entire breast instead of multiple breast locations (model M2). Both models were validated in the US Childhood Cancer Survivor Study (CCSS) cohort. We also estimated absolute BC risks for 114 Dutch and German women treated 2006-2021, and compared their model-based risks with predicted risks in the case-control study used to develop the models to assess a change in risk over time.

Results: The ERR/Gy was 0.16. Both models significantly underestimated 20-year risk in the external validation in 686 HL patients (1970-1986) from CCSS (observed/expected ratios of 1.54 for M1; 1.65 for M2), and there was no difference in discriminatory performance between models (AUC 0.68 for both). When compared to historic patients, recently treated patients received lower average breast location doses, and a smaller proportion of their breast volume received a dose of at least 10 Gy, resulting in a lower radiation-related BC risk.

Conclusion: We predicted breast cancer among HL survivors using doses to multiple locations in the breast. The discriminatory ability of the location-specific dose model was not better than using mean breast dose. Applications to other cancer sites are needed to judge the importance of accommodating dose distributions for risk prediction.

Tools for contralateral prophylactic mastectomy decision making

Mitchell H. Gail, National Cancer Institute

BACKGROUND: Women with unilateral breast cancer are increasingly opting for the removal of not only the involved breast, but also for the removal of the opposite uninvolved breast (contralateral prophylactic mastectomy [CPM]). Models to predict the absolute risk of contralateral breast cancer (CBC) can help a woman decide whether to undergo CPM. We illustrate that a better decision can be made if the patient and doctor also have estimates of the absolute risks of regional and distant recurrences and mortality from non–breast cancer causes.

METHODS: Analyses are based on two published models for CBC and published information on the hazards of regional and distant recurrences and non–breast cancer mortality. Assuming a competing risk framework and that CPM eliminates CBC but has no effect on other events, we calculate how much CPM reduces CBC risk and total risk from all these events. We propose that the benefit of CPM should reflect the reduction in CBC risk and the fraction of total risk reduced by CPM. We illustrate how these criteria affect recommendations for hypothetical women with various subtypes of breast cancer and risk factors.

RESULTS: The risk of CBC and total risk vary greatly, depending on the breast cancer subtype. In some cases, a decision for or against CPM can be based on CBC risk alone, but in others, additional consideration of total risk may cause a woman to accept or decline CPM.

CONCLUSION: There is a potential to develop more informative tools for deciding on CPM. Realizing this potential will require absolute risk models for CBC and for regional and distant recurrences and deaths from non–breast cancer based on risk factors measured before the initial surgery.

BOADICEA: a comprehensive breast and ovarian cancer risk prediction model incorporating genetic and non-genetic risk factors

Antonis Antoniou (University of Cambridge, UK)

Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, U.K.

Much more reliable and powerful cancer risk prediction be achieved by combining data on all known genetic, lifestyle and hormonal risk factors for the disease. We have recently enabled multifactorial breast and ovarian cancer risk-assessment through the BOADICEA model. This has been implemented in the CanRisk tool (www.canrisk.org) which allows healthcare professionals to obtain personalised cancer risks easily. The presentation will review the BOADICEA/CanRisk development process, the challenges in combining the effects of rare pathogenic variants in known susceptibility genes, polygenic risk scores, questionnaire-based risk factors, mammographic density and family history into multifactorial cancer risk prediction algorithms; and will review the efforts to assess the clinical validity of the predicted risks in large independent studies. The presentation will finally discuss ongoing efforts for the implementation of multifactorial cancer risk assessment in routine clinical practice for enabling cancer risk stratification and the better targeting of early detection and prevention approaches to those most likely to benefit.

Applying risk scores to practice: The challenges and pitfalls

Sam Finnikin (University Birmingham, UK)

Risk scores have been developed in a multitude of clinical areas for a variety of purposes. They may be used, for example, to set treatment thresholds, improve prognostic estimates, guide options, or inform public health strategy. In this session we will consider how risk scores can be used in individual patient encounters as part of a shared decision-making process. We will use the example of cardiovascular risk estimation to explore how risk estimation can be used in the clinical encounter and think about the different ways of presenting information.

We will also discuss the concept of shared decision making and how risk scoring fits in with this model. We will look at the literature around the utilisation of risk scoring in practice and discuss the some of the potential challenges to implementation.

Communicating empirical evidence: uncertainty, quality and trust

Dr Claudia R Schneider, University of Cambridge, UK, Department of Psychology and Winton Centre for Risk and Evidence Communication

Conveying empirical evidence is at the heart of science communication, from public health to climate change. Uncertainty, and the communication thereof, plays a central role. By nature of the scientific process, scientific information comes with uncertainties, such as around the precision of numeric estimates or the quality of the underlying evidence base. Often there is a legal or ethical imperative to communicate such uncertainties in order to inform rather than persuade, for instance in shared decision-making in medicine or informed consent. Informing includes being open about uncertainties, and presenting harms and benefits, instead of focusing unduly on one side of the story. But how do people deal with being presented with this kind of information? Can it undermine their trust in the information or the communicators? Questions pertain to both the effects of format in which empirical information is best communicated to ensure understanding and avoid unintended effects, as well as to peoples' reactions to being exposed to transparent information and uncertainties. The talk will tackle these topics and questions by presenting relevant research on the communication of evidence and uncertainty, with a focus on effects on trust.

Risk communication for disadvantaged groups

Christin Ellermann
Harding Center for Risk Literacy, University of Potsdam, DE

Informed decisions about medical treatments should be based on information that presents benefits, harms and available options in a transparent and understandable way. However, the information available often lacks essential criteria for transparent risk communication, preventing people from making informed decisions. In order to improve health care and enable informed decision-making, communication needs to involve the target audiences, taking into account their diversity. This is particularly important as certain groups in society, such as people with lower levels of education, certain age groups (e.g. older people) or people with language barriers, often have difficulties in obtaining, understanding, evaluating and using health information to make informed decisions, leading to overuse, underuse and misuse of health services and exacerbating existing health inequalities. Target group-specific information needs and preferences should already be taken into account in the planning, development, testing and evaluation of health communication.
The presentation reports on the development of evidence-based fact boxes for disadvantaged groups, which aim to contribute to increase information equity and improve health care.

Short course 3 “Clinical prediction models: development and validation”

Ewout Steyerberg[1], Ben Van Calster[1 2]

[1] Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands
[2] Department of Development and Regeneration, KU Leuven, Belgium

Clinical risk prediction models using classical or modern algorithms are everywhere in medicine. The estimated risk can be used to counsel individual patients, or decide whether or not to offer a specific intervention. In this short course we start with some general considerations for developing prediction models, such as the choice of classical or modern machine learning methods, the importance of relevant sample size and study design. We then discuss some aspects of model development, including dealing with missing values; coding of predictors; nonlinearity; model specification; and estimation. We then turn to some key issues in performance evaluation, including classic measures such as calibration and discrimination, and more novel decision-analytic concepts such as Net Benefit, in the context of external validation of prediction models. We will illustrate the concepts with medical case studies and leave ample room for interactive discussion.

Globally-accessible individual-tailored risk prediction

Prof. Donna Ankerst
Department of Mathematics, School of Computation, Information and Technology
Technical University of Munich

Six commonly used logistic regression methods for accommodating missing risk factor data from multiple heterogeneous cohorts, in which some cohorts do not collect some risk factors at all, were compared. Ten North American and European cohorts from the Prostate Biopsy Collaborative Group (PBCG) were used for fitting a risk prediction tool for clinically significant prostate cancer, defined as Gleason grade group ≥ 2 on standard TRUS prostate biopsy. External validation on a large European PBCG cohort and ten-fold leave-one-cohort-out internal validation were used to identify the optimal modeling approach based on the metrics of calibration-in-the-large (CIL), calibration curves, and area-underneath-the-receiver-operating characteristic curve (AUC). Among 12,703 biopsies from the ten training cohorts, 3,597 (28%) had clinically significant prostate cancer, compared to 1,757 of 5,540 (32%) in the external validation cohort. In external validation, the available cases method that pooled individual patient data containing all risk factors input by an end-user had the best CIL, under-predicting risks as percentages by 2.9% on average, and obtained an AUC of 75.7%. Imputation had the worst CIL (-13.3%). The available cases method was further validated as optimal in internal cross-validation and thus used for development of an online risk tool posted at riskcalc.org. For end-users of the risk tool, two risk factors were mandatory: serum prostate-specific antigen (PSA) and age, and ten were optional: digital rectal exam, prostate volume, prior negative biopsy, 5-alpha-reductase-inhibitor use, prior PSA screen, African ancestry, Hispanic ethnicity, first-degree prostate-, breast-, and second-degree prostate-cancer family history. Developers of clinical risk prediction tools should optimize use of available data and sources even in the presence of high amounts of missing data and offer options for users with missing risk factors.

Accommodating population differences in model validation

Ruth Pfeiffer, Ph.D.
Biostatistics Branch
National Cancer Institute, NIH, HHS
Bethesda, MD 20892-7244
Joint work with Yiyao Chen, Mitchell H. Gail, Donna P. Ankerst

Validation of risk prediction models in independent data provides a rigorous assessment of model performance. However, several differences between the populations that gave rise to the training and the validation data can lead to seemingly poor performance of a risk model. We formalize the notions of “similarity” of the training and validation data and define reproducibility and transportability. We address the impact of different predictor distributions and differences in verifying the outcome on model calibration, accuracy and discrimination. When individual level data from both the training and validation data sets are available, we propose and study weighted versions of the validation metrics that adjust for differences in the predictor distributions and in outcome verification to provide a more comprehensive assessment of model performance. We give conditions on the model and the training and validation populations that ensure a model's reproducibility or transportability and show how to check them. We discuss approaches to recalibrate a model. As an illustration we develop and validate a prostate cancer risk model using data from two large North American prostate cancer prevention trials, the SELECT and PLCO trials.

External validation and clinical utility assessment of the PREDICT breast cancer prognostic model in young patients with node-negative breast cancer

Yuwei Wang, MSc[1], Annegien Broeks, PhD[2], Daniele Giardiello, PhD[1, 3], Michael Hauptmann, PhD[4], Katarzyna Jóźwiak, PhD[4], Esther A. Koop, MD[5], Mark Opdam, BSc[1], Sabine Siesling, PhD[6,7], Gabe S. Sonke, MD[8], Nikolas Stathonikos, MSc[9], Natalie D. ter Hoeve, BSc[9], Elsken van der Wall, MD[10], Carolien HM. van Deurzen, MD[11], Paul J. van Diest, MD[9], Adri C. Voogd, PhD[12], Willem Vreuls, MD[13], Sabine C. Linn, MD[1, 8], Gwen MHE. Dackus, MD[1, 9§], Marjanka K. Schmidt, PhD[1, 14§]

1Department of Molecular Pathology, the Netherlands Cancer Institute, Amsterdam, the Netherlands; 2Core Facility Molecular Pathology and Biobanking, the Netherlands Cancer Institute, Amsterdam, the Netherlands; 3Eurac Research, Institute of Biomedicine, Epidemiology and Biostatistics, Bolzano, Italy; 4Institute of Biostatistics and Registry Research, Brandenburg Medical School Theodor Fontane, Neuruppin, Germany; 5Department of Pathology, Gelre Ziekenhuizen, Apeldoorn, the Netherlands; 6Department of Research and Development, Netherlands Comprehensive Cancer Organization, Utrecht, the Netherlands; 7Department of Health Technology and Services Research, Technical Medical Centre, University of Twente, Enschede, the Netherlands; 8Department of Medical Oncology, the Netherlands Cancer Institute, Amsterdam, the Netherlands; 9Department of Pathology, University Medical Center Utrecht, Utrecht, the Netherlands; 10Division of Internal Medicine and Dermatology, University Medical Center Utrecht, Utrecht, Netherlands; 11Department of Pathology, ErasmusMC Cancer Institute, Rotterdam, the Netherlands; 12Department of Epidemiology, Maastricht University, Maastricht, Netherlands; 13Department of Pathology, Canisius Wilhelmina Ziekenhuis, Nijmegen, the Netherlands; 14Department of Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands;
§ Both authors contributed equally to this work.

Background: The validity of PREDICT, a widely used breast cancer prognostic model, is unclear for young breast cancer patients. This study assessed the validity and clinical utility of the latest version of PREDICT in young, node-negative, breast cancer patients who did not receive (neo)adjuvant systemic treatment.

Methods We selected all women from the Netherlands Cancer Registry, who were diagnosed with node-negative breast cancer under age 40 between 1989 and 2000, a period in which systemic treatment was not yet deemed necessary for this patient population. Model calibration and discrimination were assessed by the ratio of observed and expected all-cause mortality (O/E ratio), and the area under the receiver-operating-characteristic-curve (AUC), respectively. Decision curve analysis was used to compare PREDICT's potential clinical utility regarding chemotherapy decision-making to a chemotherapy-to-all strategy. Patients were classified as high-risk if their predicted 10-year all-cause mortality ≥ 12% (for women with estrogen receptor [ER]-positive tumors) or ≥ 8% (for women with ER-negative tumors). Clinical utility was represented by net benefit, calculated as the rate of correctly predicted high-risk patients who should receive chemotherapy minus the weighted rate of falsely predicted high-risk patients who should not receive chemotherapy.

Results: A total of 2,263 patients with a median age at diagnosis of 36 years were included. The majority of patients had ER-positive tumors (71.1%), and 44.0% had grade 3 tumors. The median tumor size was 16mm. PREDICT significantly underestimated 10-year all-cause mortality by 33% in all patients (O/E ratio: 1.33, 95% CI: 1.22-1.43). The model discrimination was moderate overall (10-year AUC: 0.65, 95% CI: 0.62-0.67), and poor for patients with ER-negative tumors (10-year AUC: 0.56, 95% CI: 0.51, 0.61). In patients with ER-positive tumors, PREDICT showed a slightly higher net benefit of 10.0% compared to the chemotherapy-to-all strategy (net benefit: 9.8%). However, in patients with ER-negative tumors, PREDICT did not outperform the chemotherapy-to-all strategy, as both had a net benefit of 18.4%.

Conclusions: PREDICT should be used with caution in young, node-negative breast cancer patients due to its suboptimal predictive performance, especially in those with ER-negative tumors.

Methods to account for improvements in survival when developing prognostic models

Sarah Booth[1], Sarwar I. Mozumder[1,2], Lucinda Archer[3], Joie Ensor[3], Richard. D Riley[3], Paul C. Lambert[1,4], Mark J. Rutherford[1]

1 Biostatistics Research Group, Department of Population Health Sciences, University of Leicester, Leicester, UK
2 Roche Products, Welwyn Garden City, UK
3 Institute of Applied Health Research, College of Medical and Dental Sciences, University of Birmingham, UK
4 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

For many different health conditions, there have been substantial improvements in survival outcomes over time. This could be due to a variety of reasons such as the introduction of more effective treatments into clinical practice.

Prognostic models are often developed using datasets that include patients who were diagnosed across a long time period. However, although the earliest diagnosed patients have poorer survival, this is often not accounted for as part of the model development process. As a result, the prognostic model will over-estimate the risk for newly diagnosed patients.

There are a variety of methods that could be used to account for these temporal trends when developing prognostic models with time-to-event outcomes. These include modelling the year of diagnosis or applying approaches that use delayed entry techniques such as period analysis or temporal recalibration. The use of delayed entry allows a more recent subset of the data to be created which can be used to develop the model (period analysis) or to re-estimate and update the baseline of a model that does not account for trends in survival over time (temporal recalibration). This allows improvements in survival to be captured, allowing for more up-to-date predictions to be produced that are more accurate for newly diagnosed patients.

These methods can also be extended for use in a competing risk setting and this will be illustrated with an example of survival following a diagnosis of colon cancer using data from the United States Surveillance, Epidemiology and End Results (SEER) database.

When competing risks are present, risk predictions can either be produced by modelling on the subdistribution hazard scale (Fine and Gray) or by fitting multiple cause-specific hazard models. The advantages and disadvantages of each framework will be discussed.

Finally, once the required prognostic model has been developed, this raises the question of how to most effectively present and communicate the risk predictions to patients. Continuing with the colon cancer example we explore a variety of graphical approaches.

Validation and Improvement of 20 year Prediction of Cardiovascular Events with Information from Cardiac Imaging

Anika Hüsing, Nils Lehmann, Sara Schramm, Börge Schmidt, Karl-Heinz Jöckel, Andreas Stang, Raimund Erbel, on behalf of the Heinz Nixdorf Recall Study group

Institut für Medizinische Informatik, Biometrie und Epidemiologie, Universitätsklinikum Essen, Universität Duisburg-Essen

The Heinz Nixdorf Recall study (HNR) was initiated in Germany (Bochum, Essen, Mülheim) to provide long-term evidence for the predictive potential for cardiac computer tomography (CT) for cardiovascular events.

This study is based on data from 4154 participants of the HNR, who received a cardiac computer tomography (CT) at baseline between 2000 – 2003. These persons were age 45-75 without previous coronary symptoms (53% women), and were followed annually via postal or telephone questionnaires for over 20 years (median follow-up 18 years). During this follow-up time 458 participants were diagnosed with myocardial infarction or stroke; in addition 644 deaths occurred not due to cardiovascular events.

The ASCVD-(AtheroSclerotic-CardioVascular-Disease) score, based on demographic and cardiovascular risk factors was originally designed to predict 10 year risk. This score was extrapolated to 20 year risk, accounting for competing risk of death from non-cardiovascular reasons.

The degree of coronary artery calcification from baseline CT as Agatston Score (CAC-score) was added to the ASCVD risk prediction model as log-scaled continuous linear effect (ln(CAC+1)) in a Cox-regression model. Risk analysis was focussed on ASCVD risk in pre-defined categories as low, borderline, intermediate and high, and CAC-score in categories 0, 0<CAC<100, 100≤CAC<400, and 400+. Bootstrap-resampling was used to derive confidence intervals (95% CI) for measures of predictive performance.

After extrapolation and calibration the ASCVD-score was well calibrated for 20 year risk. Increased risk of cardiovascular events could be observed with increasing CAC score in men and women (together and separately) across all strata of ASCVD-risk. When CAC score was added to the ASCVD-score, Harrell's C was improved from 70.6% to 72.4% (1.9% improvement, 95%CI 1.0 - 3.0%). The net-reclassification index (NRI) of 12% (95%CI 5.3-18.1%) indicated a considerable gain in reclassification overall, which was higher in men (17%) than in women (10%). The integrated discrimination improvement (IDI) for the 20-year risk was 2.7% (95% CI 1.6- 4.2%).

In addition to the established ASCVD-score, the CT-based CAC-Score showed considerable potential to improve long-term risk-prediction of cardiovascular events over 20 years.

Reference:
Erbel R, Lehmann N, Schramm S, Schmidt B, Hüsing A, Kowall B, Hermann DM, Gronewold J, Schmermund A, Möhlenkamp S, Moebus S, Grönemeyer D, Seibel R, Stang A, Jöckel KH on behalf of the Heinz Nixdorf Recall Study Group: Diagnostic cardiac CT for the improvement of cardiovascular event prediction—twenty-year results of the Heinz Nixdorf Recall Study. Dtsch Arztebl Int 2023; 120: 25–32. DOI: 10.3238/arztebl.m2022.0360

The ethical challenges of dementia prediction

Silke Schicktanz, University Medical Center Göttingen, Dept. for Medical Ethics and History of Medicine; sschick@gwdg.de

Current research in dementia, especially on Alzheimer's disease (AD), records a shift from cure to prediction and prevention, based on a new conceptualization of dementia as a continuum. This AD continuum theory promotes a new, long phase that starts without any symptomatic changes. This stage might be detected by pathological, physiological biomarkers, 10 to 25 years before onset of the symptomatic, later stage of AD. A controversy recently emerged among ethicists and clinicians, whether such predictive information is of (any) clinical or personal value and whether such biomarkers should be offered to healthy persons or persons with subjective cognitive impairment: Should it be disclosed, and if so under which condition?

In my talk I will present major ethical issues relevant for this debate: the "right to know or not to know", the value of life planning under uncertainty and anticipated regret, as well as the fear and risk of social stigmatization. Furthermore, I will exemplarily discuss how the ethical debate hinges on issues of risk analysis, uncertainty and test validity. Finally, I will point to the general challenge, whether and how normative issues of AD predictive tests should be discussed as separate, or even prior, to empirical issues of risk information, such as test validity and quality of risk information.

How risk reduction should be expressed in primary prevention

Nicholas Wald, University College London, UK

Estimating the risk of a chronic disease in terms of the 10 year risk has several serious limitations.

- 10 year risk is arbitrary
- Risk of disease is life long
- Preventive intervention is lifelong
- Preventive intervention is inadequately defined and quantified

An improved approach will be described that overcomes these limitations.

How does the communication of scientific uncertainty affect trust in the communicator?

Charlotte Dries

Harding Center for Risk Literacy, University of Potsdam, DE

Communicating uncertainties in scientific evidence is essential, to accurately inform the public on scientific knowledge, raise public awareness of known unknowns and ensure the accountability of policy around the use of scientific evidence. However, organizations and scientists often shy away from explicitly acknowledging scientific uncertainties to the public as they fear losing trust (van der Bles et al., 2020). Is this fear warranted? So far, empirical research has provided mixed results how the communication of uncertainty affects trust in their communicators (Gustafson, 2019). One potential explanation for these mixed findings are varying contexts and audiences.

We present two studies in which we examine a specific context (change of evidence, study 1) and individual factors that may moderate the effect of uncertainty communication (study 2). In study 1 (*N*=800, convenience sample), participants read fictional information about a public health authority who announced no link between a new COVID-19 live vaccine and myocarditis. The health authority communicated either 1) no uncertainty, 2) uncertainty or uncertainty with one of two reasons for the uncertainty: 3) imprecision or 4) loss to follow-up. Participants were then informed that the health authority's statement was no longer correct as new data showed a link between the vaccine and myocarditis. Participants rated the health authority's trustworthiness before and after the evidence update. Our findings indicate that communicating uncertainty buffers against a loss in trust when evidence changes and providing an explanation for uncertainty does not harm trust. In study 2 (*N*=500, convenience sample), we set out to test different individual factors that may moderate the effect of uncertainty communication on trustworthiness perceptions, e.g. prior beliefs, preference for uncertainty communication and epistemic beliefs. The data collection is still ongoing and results will be presented at the conference.

van Der Bles, A. M., van der Linden, S., Freeman, A. L., & Spiegelhalter, D. J. (2020). The effects of communicating uncertainty on public trust in facts and numbers. *Proceedings of the National Academy of Sciences*, *117*(14), 7672-7683.

Gustafson, A., & Rice, R. E. (2020). A review of the effects of uncertainty in public science communication. *Public Understanding of Science*, *29*(6), 614-633.

Predicting the clinical course of COVID-19 patients with routine data

Martin Wolkewitz (1), Derek Hazard (1), Hamid R. Marateb (2)

(1) Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany
(2) University of Isfahan, Iran

The clinical course of COVID-19 patients can be very complex. Studies often report quantities such as the cumulative probability of admission to intensive care units, length of hospital or intensive care unit stay, duration of mechanical ventilation and mortality (1). These quantities can be estimated from routinely collected data such as dates of admission to / discharge from hospital or intensive care unit, start and stop of mechanical ventilation, transfer to other health care facilities and the vital status at discharge. In addition to the clinical follow-up data, routinely collected baseline variables such as age, sex and baseline biomarkers are often used as predictors.

In this talk, we will show the advantages of using competing risks and multistate models (2,3) for COVID-19 settings in contrast to standard models which are still frequently used in the literature. Further, we discuss following questions and issues in the COVID-19 setting: 1) how should we statistically handle the event 'transfer to other health care facilities'?, 2) are routinely collected baseline variables useful predictors for emerging variants or current waves (temporal validation)?, and 3) how do baseline predictors perform in other countries (geographical validation)?

As proposed by Spitoni et al (3), we display prediction errors based on the Brier Score as a measure of predictive accuracy that evaluates both discrimination and calibration simultaneously. We will use COVID-19 data from Freiburg (Germany), Isfahan (Iran) and Barcelona (Spain) for demonstration.

References:

1. Buttia, C., Llanaj, E., Raeisi-Dehkordi, H. et al. Prognostic models in COVID-19 infection that predict severity: a systematic review. Eur J Epidemiol 38, 355–372 (2023).
2. Spitoni, C, Lammens, V, Putter, H (2018). Prediction errors for state occupation and transition probabilities in multi-state models. Biom J, 60, 1:34-48.
3. Hazard, D, Kaier, K, von Cube, M, Grodd, M, Bugiera, L, Lambert, J, Wolkewitz, M (2020). Joint analysis of duration of ventilation, length of intensive care, and mortality of COVID-19 patients: a multistate approach. BMC Med Res Methodol, 20, 1:206.

Uncertainty in predictions

Ewout Steyerberg (Leiden University Medical Center, NL)

Pending

Performance measures for the external validation of a multi-state prediction model for the clinical progression of hospitalized COVID-19 patients

Derek Hazard, MSc

Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of

Freiburg, Freiburg, Germany

Martin Wolkewitz, PhD

Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany

Multi-state model methodology has been increasingly applied to hospitalized COVID-19 patients. These models avoid some of the most severe biases (competing risks, immortal time, selection) in hospital epidemiology and enable the rapid analysis of new data in real-time. In addition to estimating the duration of hospital stays, attention has been given to the prediction of patients' clinical courses in a multi-state setting. We developed a prediction model based on publicly available hospital data [1] from the outset of the pandemic. This training data included every hospitalized COVID-19 patient in Israel from March 1 to May 2, 2020. The model incorporated two transitory states (regular ward, intensive care unit) and two absorbing states (discharge alive, in-hospital death). The transitions among the states were adjusted for age and sex. The model was externally validated on data from 648 patients hospitalized during the initial phase of the pandemic in Freiburg, Germany.

The performance of the model was evaluated with regard to predicted state occupation and transition probabilities by calculating prediction errors using Brier and Kullback–Leibler scores as outlined in Spitoni et al.[2] . In order to account for administrative censoring, prediction errors based on inverse probability weighting and pseudo-values were implemented. Furthermore, dynamic prediction was demonstrated as a contrast to the multi-state prediction. Non-parametric Aalen Johansen estimators were calculated to determine the improvement of covariate inclusion in all models. The results will be presented as plots of the prediction error and prediction error reduction for the 4 states over 30 days after hospital admission. Results for the multi-state model showed the prediction error increases over the first 10 days for the regular ward (.25), intensive care unit (.20), and discharge alive (.25) states and then decreases steadily thereafter. The prediction error for in-hospital death increases to .04 and then plateaus until day 30.

Consideration was also given on how to model changing circumstances of the pandemic (new treatments, variants, vaccinations, etc.) via surrogate measures (e.g. time since pandemic outset). The development of summary measures for the prediction model was also explored. External validations on data from three additional sites are in planning.

[1] Roimi, Michael, et al. "Development and validation of a machine learning model predicting illness trajectory and hospital utilization of COVID-19 patients: a nationwide study." *Journal of the American Medical Informatics Association* 28.6 (2021): 1188-1196.

[2] Spitoni, Cristian, Violette Lammens, and Hein Putter. "Prediction errors for state occupation and transition probabilities in multi-state models." Biometrical Journal 60.1 (2018): 34-48.

Effective sample size: expressing individual uncertainty in predictions

Doranne Thomassen1, Saskia le Cessie1, Hans van Houwelingen1, Ewout Steyerberg1

1 Dept. of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands

With healthcare becoming centered on the individual patient, individual (risk) predictions play an increasingly important role. When a clinical prediction model is developed, not all types of patients are represented equally in the observed data. As a result, epistemic uncertainty about individual risk predictions may vary widely between patients. Our aim was to develop an intuitive concept to assess and communicate uncertainty about individual risk predictions: the effective sample size.

For a given patient, the variance of their predicted risk can be equated to the variance of the sample mean outcome in n* hypothetical patients with the same model parameter values. This hypothetical sample size n* can be interpreted as the effective sample size of similar patients in the data that informed the prediction model. Similarity is model-dependent. Assuming that the prediction model is correctly specified, the concept of effective sample size can be used to express certainty about predictions to patients in terms of a sample size; for instance, by communicating that “this prediction is effectively based on 5 people like you.”

We have derived analytical expressions to calculate effective sample sizes for prediction models that are based on linear or logistic regression, or any other generalized linear model. For machine learning models or other complex models, bootstrap resampling could be applied to translate the standard error of a prediction into an effective sample size. The concept of effective sample size was illustrated in a large clinical dataset (n=1216) of patients with myocardial infarction. We found large differences in effective sample sizes between patients.

In sum, we propose translating the standard error of a prediction into an effective sample size, which could serve as an intuitive measure of uncertainty in individual predictions. Empirical research is required to determine the value of this presentation of uncertainty to patients in a shared decision-making process.

Clinical utility of risk stratification for cancer screening at the population level

Nora Pashayan, University College London, UK

Risk assessment per se does not have inherent clinical utility; the subsequent adoption of a risk-based intervention based on the results of the assessment is what influences the health outcomes. The use of such a strategy depends on whether the risk-based intervention is appropriate, accessible, practicable and acceptable. To demonstrate these, the benefit-harm balance and cost-effectiveness of risk-stratified prostate and breast cancer screening strategies and implementation considerations will be discussed.

Recalibrating Risk Models for Maximum Net Benefit

Kathleen Kerr, University of Washington, US

A risk model is miscalibrated if predicted risks do not accurately capture event rates. Sometimes it is possible to identify and address the cause of miscalibration, or build a new risk model to replace the miscalibrated model. In other circumstances, it may be necessary or desirable to recalibrate the existing risk model.

Most recalibration methods are generic and do not account for how the risk model will be used. However, our interest is settings in which the risk model will be used for risk-based clinical decision-making and standardized net benefit is the measure of risk model performance. We propose new recalibration methods that, directly or indirectly, prioritize good calibration around the critical risk threshold where good calibration is most important and affects clinical decision-making. The new methods are parsimonious and extensions of Cox's logistic recalibration. We also propose a graphical tool for assessing the potential for recalibration to improve the net benefit of a risk model.

REFERENCE: Mishra A, McClelland RL, Inoue LYT, Kerr KF. Recalibration Methods for Improved Clinical Utility of Risk Scores. Medical Decision Making, 2022. All authors (Anu Mishra, Robyn L. McClelland, Lurdes Y.T. Inoue, Kathleen F. Kerr) were members of the Department of Biostatistics at the University of Washington at the time of this research. Dr. Mishra's current affiliation is Imperial College London.

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

M.A.E. Binuya[1,2,3], A.H. Boekhout[4], S.C. Linn[5,6,7], E.G. Engelhardt[1,4], M.K. Schmidt[1,3]

[1] Division of Molecular Pathology, the Netherlands Cancer Institute – Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands
[2] Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands
[3] Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands
[4] Division of Psychosocial Research and Epidemiology, the Netherlands Cancer Institute – Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands
[5] Department of Molecular Pathology, the Netherlands Cancer Institute – Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands
[6] Department of Medical Oncology, the Netherlands Cancer Institute – Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands
[7] Department of Pathology, University Medical Center Utrecht, Utrecht, The Netherlands

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.

Family decisions in genomic newborn screening: Subproject Medical Psychology in the NEW_LIVES project

Elena Sophia Doll[1,*], Seraina Petra Lerch[1], Julia Mahal[1], Beate Ditzen[1]

*ElenaSophia.Doll@med.uni-heidelberg.de

[1]Institute of Medical Psychology, Heidelberg University Hospital, Ruprecht-Karls University Heidelberg, Heidelberg, Germany

Background: For an increasing number of diseases and disease predispositions, genetic causes have been identified. This provides individuals and families with the possibility of learning more about hereditary predictors and genetic susceptibility for known disorders or clinical manifestations, which would otherwise remain undetermined. At the same time, improved diagnostic methods in genetic medicine—namely next generation sequencing (NGS)—make it possible to simultaneously identify multiple genes associated with different diseases that might not (yet) have resulted in any symptoms or might not even affect the person tested, but may have implications for his or her children. Based on this, different countries now evaluate scenarios for genomic screening of (yet) asymptomatic newborns (gNBS). This development results in far-reaching consequences for risk perception and decision processes for the parents and families facing these test opportunities. However, psychological data on gNBS are still sparse and research is predominantly from North America.

Aim: The psychological part of the BMBF-funded project NEW_LIVES (NEWborn screening programs – Legal implications, Values, Ethics and Society) aims to identify the opinions and needs of different stakeholders and parents' decision making processes in the context of gNBS.

Methodology: Currently, we are preparing a review on relevant factors for parents' decision making in pediatric genetic testing. Additionally, we conceptualized focus groups with parents, adult patients with a genetic disorder, patient representatives, and healthcare professionals to obtain an overview of perceived opportunities and risks regarding gNBS as well as requirements for the information and consent process. Based on this, we will develop case scenarios and a value clarification exercise to be used in an online survey to relate assessments and preferences about gNBS to standardized questionnaire data (e.g., decision conflict, values, risk perception).

Expected Results: No results are available yet. Expected outcomes include parents' desire to be informed about gNBS results depending on actionably, but also their own medical history, values, and family structure.

Expected benefits and outlook: The results of our research will inform on decision processes and decision aids in genetic medicine and gNBS in particular.

Keywords: Genetic testing, genomic newborn screening, decision making, risk perception

Fluid-biomarker based Prediction of Alzheimer's disease

Anja Schneider, German Center for Neurodegenerative Diseases (DZNE), DE

Alzheimer's disease (AD) is increasingly recognized as a disease continuum which starts decades before clinical symptoms become manifest. New disease-modifying treatments are supposed to be most effective when started at very early disease stages, ideally even in the preclinical, largely asymptomatic disease stage. This need has accelerated biomarker research to develop low invasive markers for the prognosis and diagnosis of Alzheimer's disease pathology, primarily intended in the context of research applications. With the advent of the first approved amyloid-targeted disease-modifying drugs, blood-based biomarkers are widely discussed for case-finding purposes in primary care physician settings.

Here, I will provide an overview on biomarker-based classification of Alzheimer's disease, new developments in AD blood biomarker research, and critically acknowledge limitations and requirements for their implementation in routine clinical care.

Ethics of prediction in neurodegenerative diseases

Ineke Bolt, Max Rensink, Maartje Schermer

Department of Medical Ethics, Philosophy & History of Medicine, Erasmus Medical Center, Rotterdam, The Netherlands

Predictive genetic testing is available to test whether individuals at risk for neurodegenerative diseases (e.g., Huntington's disease) carry the mutated gene. So far, predictive genetic tests cannot predict the age of onset (AO) and the severity or progression of disease (POD). Currently, new biomarkers are being developed in order to predict the age of onset (AO), severity, and progression of disease symptoms (POD) of genetic neurodegenerative diseases, including Huntington Disease (HD), Spinocerebellar Ataxia (SCA), and Frontotemporal Dementia (FTD). Onset prediction testing offers gene carriers a new opportunity, that is to learn *when* they will develop the first symptoms and/or how *severe* the symptoms will be. The value of these tests can be threefold: 1. for personal use by potential gene carriers in life decision-making: at-risk individuals can use this information to make decisions about future life plans and reproduction, 2. in clinical research settings: accurate prediction is needed for participation in clinical trials, and 3. in clinical care settings: to estimate the exact timing to start medication – assuming that, in the future, medical treatment becomes available.

The aim of this presentation is to identify the ethical issues raised by AO and POD prediction for neurodegenerative diseases in research settings as well as in clinical practice and to provide a research agenda. In order to identify these issues, first a short overview of the ethical issues of current predictive genetic testing of neurodegenerative disease (With Huntington's disease as a model since the 1980s) will be provided. Next, the ethical issues raised by biomarker testing for multifactorial neurodegenerative diseases, in particular Alzheimer's disease, will be reviewed. Since biomarker testing for multifactorial neurodegenerative diseases may be comparable to AO and POD prediction, it can help us to identify the issues of AO and POD prediction. Finally, a research agenda is formulated including the ethical issues that need to be addressed for a responsible research-setting and implementation of AO and POD prediction. Amongst the topics discussed are complexity of risk information & informed consent, accuracy of Artificial Intelligence-generated AO and POD prediction, and implications of disease definition and classification.

Prognostic Prediction in Relapsing-Remitting Multiple Sclerosis – the way forward?

Begum Irmak On, Anna Maria Sakr, Ulrich Mansmann

Institute for Medical Information Processing, Biometry and Epidemiology, Ludwig-Maximilians-Universität München, Munich, Germany

The clinicians in the field of relapsing-remitting multiple sclerosis (RRMS), a multifactorial and chronic disease, describe its course as heterogeneous and unpredictable. The recent availability of over a dozen treatment options with various safety and efficacy profiles makes individualized risk prediction an aspired goal (1). In a collaborative project, we conducted a systematic review (2) to identify and evaluate the quality of existing prognostic prediction models for common clinical endpoints. Of the 75 models included, 40 were based on machine learning methods, 12 were externally validated, and 73 had high risk of bias in their analysis or external validation, probably leading to inflated performance. The methodologically sound models predicting 2-year disability had area under the curve (AUC) of 0.59 and 0.66 in internal validation and were not fully reported to allow for their independent validation.

In addition to the review, we used datasets from different repositories to explore the potential of prediction for RRMS patients applying state-of-the-art strategies. We developed a transformation forest model to predict 2-year cerebral lesions using routine care data (DIFUTURE), Bayesian models to predict relapses and disability using registry data (OFSEP), and elastic net to predict the above-mentioned efficacy endpoints at 2-years using trial data (CSDR). The cross- or externally validated AUC of these models varied between 0.59 and 0.74 and tree-based methods were not superior to regression models. Our findings and interactions with medical colleagues reveal a fixation on obtaining a high AUC or accuracy. Also, there is underappreciation of the potential usefulness of well-developed models with moderate discrimination, and the need for further validation and impact studies. There is the expectation that model performance will increase with novel machine learning methods, or by using predictors and outcomes that are uncommon, expensive, and difficult to collect or analyze. These lead to further research waste due to the low quality or applicability of the developed models. The current challenge is engaging with the medical community to increase awareness about how to evaluate the methods and interpret the results from prognostic prediction research, and which qualities a prediction model needs to possess before its clinical implementation.

1. Thompson AJ, Baranzini SE, Geurts J, Hemmer B, Ciccarelli O. Multiple sclerosis. Lancet. 2018;391(10130):1622-1636. doi:10.1016/S0140-6736(18)30481-1
2. Seker BIO, Reeve K, Havla J, et al. Prognostic models for predicting clinical disease progression, worsening and activity in people with multiple sclerosis. Cochrane Database of Systematic Reviews. 2020;(5). doi:10.1002/14651858.CD013606

Peak performance: does a simple statistician understand random forests for risk prediction?

Ben van Calster (KU Leuven, BE)

Random forests have become popular for clinical risk prediction modeling. In a case study on predicting ovarian malignancy, we observed training AUCs close to 1. Although this suggests overfitting, performance was competitive on test data. We aimed to understand the behavior of random forests by visualizing data space for the case study. Visualization of data space suggested that the model learned 'peaks of probability' around training set events. A cluster of events created a big peak (signal), isolated events local peaks (noise).

We then performed a simulation study with 192 scenarios and 1000 simulation runs. In short, we observed near perfect median training AUCs except in scenarios with only a few binary predictors, or scenarios with many binary predictors and high minimum node size. Median test AUCs were higher with higher events per variable, higher minimum node size, and binary predictors. Median training calibration slopes were always >1. Median test slopes ranged between 0.45 and 2.34, and were not related to median training slopes. Median test slopes were higher with higher true c-statistic, higher minimum node size, and higher sample size.

We conclude that random forests learn local probability peaks, often yielding near perfect training AUCs. The simulation results disagree with the common recommendation to use fully grown trees, and suggest that calibration performance is erratic.

Decreasing complexity of risk prediction models by introducing Discriminative Power Lasso

Cornelia Fuetterer (Technical University of Munich (TUM), TUM School of Medicine,
Institute of AI and Informatics in Medicine)

For personalized treatment decisions, risk prediction models are of high importance especially in cancer research. For instance in breast cancer research, specific risk factors are known for specific subpopulations, such as certain clinical covariates as well as genetic covariates. The main goal of the according prediction models is to achieve good calibration, predicting well the observed outcome, as well as a high discrimination. With the regularized regression of the least absolute shrinkage and selection operator (Lasso) we aim to identify a sparse set of covariates that can be used for prediction, selecting the most impactful covariates. We propose an adapted penalization aiming at a better discrimination of high and low risk patients based on the fact that decisive covariates are differently distributed among these subpopulations. We measure these differences by their discriminative power (DP), which includes univariate compactness within classes and separation between classes. The construction of the covariate specific DP measures include concepts of ANOVA as well as of clustering theory. The DPs are then integrated as covariate specific discount factors into the penalization term of the original adaptive Lasso, such that covariates with a higher DP are penalized less and thus have a higher chance of remaining in the final model.
The resulting model, that we call Discriminative Power Lasso (DP-Lasso) aims to increase the discrimination of the model. We thus provide the selection of more promising and trustworthy covariates, while the coefficients of uninformative covariates can be shrunken to zero more reliably. We test our method on genetic data as well as on simulated data. DP-Lasso leads on average to considerably sparser solutions compared to competing Lasso-based regularization approaches, while being competitive in terms of accuracy.

Keywords – Penalized Regression, Variable Selection, Clustering validation metrics

How to statistically model biologic interactions

Carolin Malsch, Institute for Mathematics and Informatics, University of Greifswald, Germany

After decades of statistical modeling, there is still no clear concept how to assess interaction effects of a set of binary factors on a binary response variable. The reason for this seems to be a lack of clarity about how absence of biologic interaction is modeled.

Biological interaction between two risk factors is often understood as either a deviation from additivity of the absolute effects of two (or more) factors, or non-zero coefficients for interaction terms in binary logistic regression. Both approaches are incorrect.

The mathematically adequate concept for modeling biological (non-)interaction in the given context is stochastic (in-)dependence. Hence, strategies and software recommendations provided in the literature to date are misleading and need correction.

Affected by this misunderstanding is also how logistic regression analysis, the most common approach to model the joint effect of two or more factors on a binary response variable in health research, is conducted in application. In most cases, only main effects are estimated in the regression function while interaction terms are omitted completely. Only sometimes a selection of interaction terms is taken into account with the aim to assess biologic interaction.

In the binary logistic regression model, interaction terms do not reflect biological interactions in general. For example, they are inevitably needed to model stochastic, and thus biologic, independence. On the other hand, coefficients of interaction terms take on value zero when a special type of stochastic dependence is present.

Missing out on interaction terms in the logistic regression model leads to severely biased estimates and easily causes misleading interpretation. This is particularly worrying given that results from studies in epidemiology, health services and public health eventually affect clinical and public health recommendations.

To resolve these problems, this contribution seeks to clarify (a) how biologic interactions in are correctly assessed using stochastic (in-)dependence, (b) why interaction terms in a binary regression model must be considered in the regression function and (c) which value they take on in case of absence of biologic interaction.

The related theory is presented and demonstrated on examples. Further, other approaches to assess biologic interactions from data are critically discussed.

Key words: binary logistic regression, biologic interaction, epidemiology, public health, stochastic dependence