

University of Groningen

Contemporary issues in static and dynamic prediction

Chen, Yuntao

DOI:
[10.33612/diss.232289260](https://doi.org/10.33612/diss.232289260)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2022

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Chen, Y. (2022). *Contemporary issues in static and dynamic prediction: some applications and evaluation in the clinical context*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. <https://doi.org/10.33612/diss.232289260>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Chapter 8
Nederlandse Samenvatting; About the author;
Acknowledgements

NEDERLANDSE SAMENVATTING

Voorspellingsmodellen die de kans op het ontwikkelen van een specifieke ziekte (diagnostisch model) of een specifiek eindpunt van de ziekte (prognostisch model) schatten, gegeven een reeks kenmerken van de proefpersoon, zijn nauw verbonden met gepersonaliseerde geneeskunde waarvan het kernidee is om medische beslissingen te baseren op de individuele patiënt kenmerken in plaats van op populatiegemiddelden. Afhankelijk van het beslispunt kunnen voorspellingsmodellen worden onderverdeeld in twee categorieën: statische voorspellingsmodellen (eenmalig een beslissing nemen) en dynamische voorspellingsmodellen (dynamisch bijgewerkte beslissingen nemen). Hoewel multivariabele logistische en Cox-regressie vaak worden gebruikt om voorspellingsmodellen te ontwikkelen, zijn ze niet de hoofdsleutel voor elke situatie. Verschillende problemen zoals geclusterde gegevens, concurrerende risico's en tijdsafhankelijke variabelen kunnen optreden wanneer een eenvoudig logistisch of Cox-model het risico niet correct kan inschatten in statische en dynamische voorspelling. Hoewel er aangepaste of meer geavanceerde benaderingen zijn ontwikkeld om deze problemen op het gebied van medische statistiek aan te pakken, worden ze niet op de juiste manier toegepast in klinisch onderzoek. Om deze leemte op te vullen, illustreerde dit proefschrift hoe geavanceerde statistische modellen op de juiste manier kunnen worden toegepast om betere voorspellingen te verkrijgen met behulp van een reeks klinische casestudies.

Het eerste deel van het proefschrift bestaat uit twee studies die zich richten op de kwestie van heterogeniteit tussen studies en concurrerende risico's bij statische voorspelling. Hoofdstuk 21 presenteerde een praktisch voorbeeld van het ontwikkelen van een klinisch voorspellingsmodel door de heterogeniteit tussen studies in de populatie aan te pakken. Het doel van deze studie was om een gestratificeerd model voor hartfalen (HF) fenotype te ontwikkelen voor het voorspellen van 1-jaars mortaliteit bij patiënten opgenomen met acuut HF door middel van een meta-analyse van individuele deelnemersgegevens van vier Europese cohorten. In deze studie kwamen twee aspecten van heterogeniteit aan de orde. Een daarvan is de mogelijk verschillende baseline sterftecijfers van drie HF-subtypes. De andere is de mogelijk verschillende voorspeller-uitkomst-associaties tussen HF-fenotypen. Met behulp van gegevens verzameld van 3577 patiënten in vier Europese cohorten, ontwikkelden we een HF-fenotype gestratificeerd model met 11 gemakkelijk beschikbare voorspellers. Vier van de voorspellers, namelijk systolische bloeddruk, serumcreatinine, myocardinfarct en diabetes, beïnvloedden het sterfterisico verschillend in de HF-fenotypes. Het model liet een uitstekend onderscheidingsvermogen zien met een gewogen intern-externe kruisgevalideerde AUC van 0,79 (0,74-0,83) voor HF met verminderde ejection fraction, 0,74 (0,70-0,79) voor HF met middelhoge ejection fraction en 0,74 (0,71-0,77) voor HF met bewaarde ejection fraction. Kalibratie was ook goed met de gemiddelde voorspelde 1-jaars sterfterisico's dicht bij de gemiddeld waargenomen risico's, vooral na herijking van de baseline sterfterisico's. Ons model toonde uitstekende generaliseerbaarheid over

vier Europese cohorten en kan een nuttig hulpmiddel zijn bij HF fenotype-specifieke klinische besluitvorming. In Hoofdstuk 32 gebruikten we gegevens van een retrospectief cohort in Wuhan, China om een prognostisch model te ontwikkelen om ziekenhuissterfte bij COVID-19-patiënten te voorspellen. In deze studie werd een concurrerende risicoanalyse uitgevoerd om levend ontslagen als concurrerend risico te behandelen om overschatting van de mortaliteit in het ziekenhuis te voorkomen. Het uiteindelijke model met vijf routinematig gemeten demografische en klinische voorspellers werd intern gevalideerd met behulp van de bootstrap-benadering en extern gevalideerd in een ander cohort in Wuhan. Het model vertoonde een zeer goede discriminerende en kalibratienauwkeurigheid. Het model kan artsen helpen om de patiënten vroegtijdig te stratificeren op basis van de geschatte mortaliteit op 7 dagen (14 dagen of 30 dagen) na opname, waardoor patiënten gerichte ondersteunende zorg krijgen en de beperkte medische voorzieningen (bijv. beademingsapparatuur) beter worden toegewezen. Vooral wanneer de capaciteit van de kritieke zorg overbelast is.

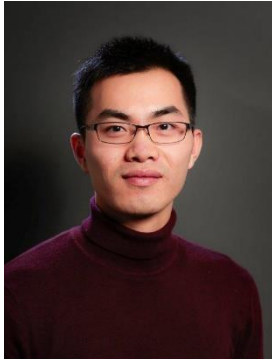
Het tweede deel van het proefschrift omvat drie studies met betrekking tot dynamische voorspelling. Hoofdstuk 4 en 5 presenteren de toepassing van twee verschillende gezamenlijke modelleringsbenaderingen in de klinische setting, en illustreerden hoe we een tijdsafhankelijke variabele konden modelleren en de associatie ervan met een overlevingsresultaat konden bestuderen. Hoofdstuk 6 evalueerde empirisch de voorspellende nauwkeurigheid van de hierboven geïntroduceerde gezamenlijke modellen in een klinische context. In hoofdstuk 43

hebben we specifiek het shared random effects model (SREM) geïntroduceerd om de associatie tussen een continue tijdsafhankelijke variabele en een overlevingsresultaat te bestuderen met behulp van een voorbeeld in de respiratoire geneeskunde. We hebben lezers kennis laten maken met de terminologie van de SREM en uitgelegd waarom, in vergelijking met het meer algemeen toegepaste tijdsafhankelijke Cox-model (TDCM), de SREM naar verwachting zal resulteren in minder vertekende schattingen van het effect van de tijdsafhankelijke variabele. Een praktisch voorbeeld dat de associatie tussen herhaalde metingen van de centrale apneu-hypopneu-index (cAHI) en cardiovasculaire mortaliteit bestudeert, werd gebruikt om de aanpak te illustreren. Zowel de SREM als de TDCM werden gefit om de associatie tussen de huidige waarde van cAHI en het risico op cardiovasculaire mortaliteit te schatten. We vonden dat de geschatte HR van de TDCM werd verzwakt zoals verwacht op basis van de statistische literatuur. Onze studie toonde aan dat de SREM een geschiktere benadering kan zijn om de associatie tussen een continue tijdsafhankelijke variabele en een overlevingsresultaat te bestuderen. Hoofdstuk 5 presenteerde een toepassing van een misclassificatiemodel, een belangrijk speciaal geval van het Hidden Markov-model (HMM), om de associatie tussen een categorische tijdsafhankelijke variabele en een overlevingsresultaat te bestuderen. Ons doel was om de associatie tussen het stadium van chronische nierziekte (CKD) en nieuw optredende HF te bestuderen door rekening te houden met afnemende nierfunctie met de leeftijd en mogelijke misclassificatie van CKD-stadia. Eerdere studies hebben aangetoond dat nierdisfunctie geassocieerd is met incident HF⁴. Er is

echter minder bekend over hoe het risico van een proefpersoon om HF te ontwikkelen verandert met de leeftijd als reactie op de verandering van de proefpersoon in het CKD-stadium met de leeftijd. Om deze vraag te beantwoorden, hebben we een multi-state model uitgerust met drie CKD-stadia als voorbijgaande toestanden en nieuw optredende HF (uitkomst van interesse) en overlijden vóór nieuw optredend HF (concurrerend risico) als absorberende toestanden. We vonden dat de risico-associatie tussen CKD-stadium en nieuw optredende HF op de relatieve schaal verzwakte met de leeftijd, maar sterker werd met de leeftijd op de absolute schaal. Bovendien waren de geschatte incidentiecijfers tussen toestanden (bijv. minder ernstig CKD-stadium tot ernstiger CKD-stadium en verschillende CKD-stadia tot HF/dood) zeer informatief bij het beschrijven van het individuele risico van overgang van de ene toestand naar de andere, waardoor de cardiorenale ziekte werd afgebakend. De SREM die in Hoofdstuk 4 is geïntroduceerd, en de HMM die in Hoofdstuk 5 wordt gebruikt, kunnen beide worden gebruikt om de ziekteprognose dynamisch te voorspellen. In Hoofdstuk 6 werden de voorspellende prestaties van de twee modellen empirisch vergeleken in de context van het dynamisch voorspellen van mortaliteit door alle oorzaken voor patiënten met acuut HF op basis van seriële NT-proBNP metingen. Onze resultaten toonden aan dat dynamische voorspellingsbenaderingen die seriële NT-proBNP-metingen bevatten, het potentieel hebben om de voorspellingsnauwkeurigheid te verbeteren in vergelijking met de benadering waarbij alleen basislijn-NT-proBNP-metingen worden gebruikt. De HMM presteerde beter dan de SREM bij het dynamisch voorspellen van mortaliteit

met behulp van seriële NT-proBNP-metingen in dit patiëntencohort. Daarom kan de HMM een nuttig hulpmiddel zijn om dynamische voorspelling in klinisch onderzoek te ondersteunen.

ABOUT THE AUTHOR



Yuntao (Simon) Chen was born in a small but beautiful city Yixing, Jiangsu on October 4, 1990. He initially studied Preventive Medicine in the School of Public Health, Nanjing Medical University. After 5 years of medical training, he then found an interest in medical statistics and was enrolled in Fudan University to pursue a Master's degree in biostatistics in 2014. During the master's study, he focused on using electronic health records to study the association between repeated biomarker measurements and prognosis of pancreatic cancer. In 2017, he came to the Netherlands and started his PhD in medical statistics at University Medical Center Groningen. During the four years PhD trajectory, He has been supervised by several exceptional researchers including, but not limited to Professor Gerton Lunter, Dr. Douwe Postmus, Professor Christine Eulenburg and Professor Marike Boezen. He has worked on various projects with the aim to bridge the gap between statistical and clinical research. The primary focus lies in how advanced statistical approaches can improve predictive accuracy in static prediction models that use baseline information only and dynamic prediction models that can additionally incorporate follow-up information.

Since October 2021, he has been working as a postdoctoral research fellow mentored by Professor Eric Brunner in the Department of Epidemiology and Public Health at UCL. He will mainly focus on using multi-state model to predict and compare future dementia burden in UK and China. Besides, he will explore the association of midlife cardiac biomarkers with cognitive impairment and dementia using the WHITEHALL II study.

After 5 years of medical training and 8 years of research training, Yuntao Chen finds that there remains a huge gap between research and clinical practice. Regarding the research area, from researching cancer in Master, to researching heart failure in PhD, and finally to dementia in the postdoc, he decides to devote himself to the aging field. The reason is simple: the aging process is unavoidable and some diseases such as dementia are currently untreatable, which emphasizes the importance of preventive measures. Taking advantage of all the experiences, he is willing to use state-of-art statistical approaches to study the role of shared risk factors such as midlife hypertension in the development of cardiovascular disease and dementia.

PUBLICATIONS

Main research

1. **Chen Y**, Postmus D, Cowie M R, et al. Using joint modelling to assess the association between a time-varying biomarker and a survival outcome: an illustrative example in respiratory medicine. *European Respiratory Journal*, 2020. DOI: 10.1183/13993003.03206-2020.
2. **Chen Y**, Voors AA, Jaarsma T, et al. A heart failure phenotype stratified model for predicting 1-year mortality in patients admitted with acute heart failure: results from an individual participant data meta-analysis of four prospective European cohorts. *BMC Med*. 2021 Jan 27;19(1):21. DOI: 10.1186/s12916-020-01894-2.
3. Li J*, **Chen Y***, Chen S, et al. Derivation and validation of a prognostic model for predicting in-hospital mortality in patients admitted with COVID-19 in Wuhan, China: the PLANS (platelet lymphocyte age neutrophil sex) model. *BMC Infect Dis*. 2020 Dec 17;20(1):959. DOI: 10.1186/s12879-020-05688-y.

Collaborations with clinical researchers

1. Li A, **Chen Y**, van der Sluis LWM, Schuller AA, Tjakkes GH. Mediation Analysis of White Blood Cell Count on the Association Between Periodontal Inflammation and Digit Symbol Substitution Test Scoring Cognitive Function Among Older U.S. Adults. *The Journals of gerontology. Series A, Biological Sciences and Medical Sciences*. 2020 Sep. DOI: 10.1093/gerona/glaa223.
2. Li A*, **Chen Y***, Schuller AA, van der Sluis LWM, Tjakkes GE. Dietary Inflammatory Potential is Associated with Poor Periodontal Health: A Population-based Study. *J Clin Periodontol*. 2021 Apr 25. DOI: 10.1111/jcpe.13472.
3. Yang X, Chen D, **Chen Y**, Wang N, Lyv C, Li Y, Jie J, Zhou T, Li Y, Zhou P. Geographical distribution and prevalence of mental disorders among healthcare workers in China: A cross-sectional country-wide survey. *Int J Health Plann Manage*. 2021 May 8. DOI: 10.1002/hpm.3186.
4. Li A, Vermaire JH, **Chen Y**, van der Sluis LWM, Thomas RZ, Tjakkes GE, Schuller AA. Trends in socioeconomic inequality of periodontal health status among Dutch adults: a repeated cross-sectional analysis over two decades. *BMC Oral Health*. 2021 Jul 15;21(1):346. DOI: 10.1186/s12903-021-01713-x.
5. Li A*, **Chen Y***, Visser A, Marks LAM, Tjakkes GE. Combined association of cognitive impairment and poor oral health on mortality risk in older adults: Results from the NHANES with 15 years of follow-up. *J Periodontol*. 2022 Jun;93(6):888-900. DOI: 10.1002/JPER.21-0292.

*shared first authors

ACKNOWLEDGEMENTS

Finally, it comes to this stage. If you ask me whether it is worthwhile to take 1/6 of the years I lived till now, and possibly 1/15 of my lifetime (assuming I could live to 75) to pursue a PhD, my answer is “It is!” and “absolutely it is given that you are doing a PhD in the Netherlands”. The PhD life in the past 4 years in the Netherlands has been an eye-opening and rewarding experience for me.

First and foremost, I will narrate my PhD story and thank all the supervisors (Prof. Christine Eulenburg, Dr. Douwe Postmus, Prof. Marike Boezen and Prof. Gerton Lunter) along the way. The PhD trajectory was difficult at some period. I conquered it in the end which would not be possible without any of you.

Dear Christine, you brought me to Groningen four years ago. After several emails regarding my PhD proposal, I was immediately attracted by your German-style rigour. You did not guarantee the PhD position until around six rounds of revision of my PhD proposal. After I started the PhD, the rigour continued. One example was that I repeatedly forgot to add page number in slides and you repeatedly reminded me of the importance of adding it. You introduced Dr Douwe Postmus to my PhD trajectory as the co-promotor. This later proves to be the most important and accurate decision. You left in my second year because of personal reasons. However, I would like to thank you for giving me the chance to study here and helping me settle down in the department.

Dear Douwe, my first impression of you is tall and stern (when you do not smile). After talking more with you, I find you are super easygoing and open-minded. Academically, you are a real scientist and what is even more commendable is that you are always approachable. Whenever I encounter any issues that can be related to as large as study design or as small as coding details, I can directly go to your office without any appointments. I still remember that we often discuss coding issues for more than two hours during my first two years when I did not code well enough. I also remember that we usually start with a superficial question, and gradually become a clear picture on the whiteboard regarding what I should do in the next step. You walk me through every aspect of research, which I am sure will benefit my entire research career. Thank you so much, Douwe, sincerely. I hope we can occasionally have a drink in Groningen, London, Shanghai or somewhere else. That would be great.

I am also missing Prof. Marike Boezen so much. You supervised me before Gerton took over. With fancy suits, you seem stern and unapproachable. However, my impression of you is humorous, confident and generous after having several relaxing chats with you over several serious topics. May you rest in peace!

Dear Gerton, thank you so much for supporting me through the last part of my PhD trajectory. I know it is not easy to take over a PhD student halfway with a project that does not match with your expertise. However, you are very supportive and responsible. Trained in math, you are rigorous about my writing regarding statistical

assumptions. I appreciate this so much since it gives me a deeper understanding about those assumptions. You are a good leader to make our medical statistics unit a better team. You have organized statistical meetings encouraging us to present our work. You have also organized coffee time to chill in the park during the pandemic. You also gave me a lot of suggestions about my career development. I still remember that you told me that the principle of a good job is that the team benefits from you and you also benefit from the team.

Dear Prof. Hans Hillege, maybe I should list you as my supervisor as well. Thank you so much for giving me access to most of the data I used in my PhD. Together with Douwe and Gerton, you supervised me over the work related to cardiorenal syndrome. Your knowledge in this field is enormous. I also enjoyed the talk with you regarding career development and work-life balance.

I would like to thank all the co-authors and collaborators who have worked with me and contributed to this thesis. I have learned a lot from you and hope I have the chance to collaborate with you in the future.

To the members of the assessment committee: Prof. G.H. de Bock from University Medical Center Groningen, Prof. O.H. Klungel from Utrecht University and Prof. P. van der Harst from University Medical Center Utrecht, many thanks for reading my thesis and giving valuable suggestions.

Yi and Lili, many thanks for being my paranymphs. Yi, you are the only person that will make ‘Bimodal Forum’ shudder on the table. I hope you will also have an unforgettable defense ceremony several weeks later. Lili, you seem to be that kind of person belonging to academics. And you truly are. We had many constructive discussions about our projects. I am always feeling that you have comprehensive epidemiological thinking over many topics.

Big thanks to all my colleagues and friends in the Department of Epidemiology. Tian, Yihui, Jing Wang, Du Jing, and Qing, thank you for all the accompanies along the entire PhD trajectory. At the very beginning, we helped each other and settled into Groningen life. I enjoyed all the parties and outings with you. Dear “Cool Epi” lunch group members: Elnaz, Bale, Nigus, Kebede, Sisay, Mele, Pato, Omar, Aruka, Rujia, Katri, Tian, I very much enjoyed all the lunches, parties, football matches and dances with you. I hope you all the best in work and life. Dear Reinder, Tesfa, Thomas, Ahmad, Tugs, Senshuang, Jiali, Xueling, Rujia, Zekai, Wenbo and many others, thank you all for the accompany. I wish you all a bright future.

Dear Aukje, Roelian, Lisette and Erwin, thanks a lot for supporting me with all those administrative and financial things.

To all the other Chinese friends I met in Groningen, Qi, Yichen, Cancan, Yanni, Miaozhen, Yuanyuan, Yue, Boyan, Sunyi and many others, I really enjoyed all the talks and parties with you, especially in Groningen’s bleak winter. I wish you all the best.

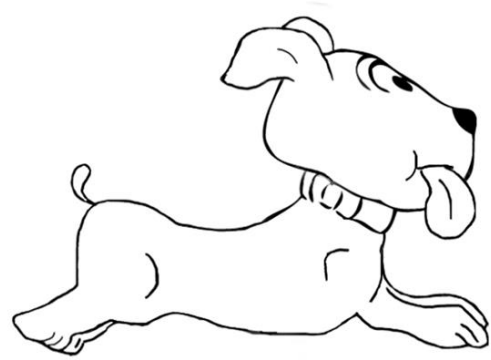
To committee members of ‘Bimodal Forum’, Mengfei Cai, Lianmin Chen, Shixain Hu, An Li and Bingjiang Qiu. My defense seems to be the final section of the Dutch chapter of this Forum. However, I know it has already begun its Chinese

chapter, marked by Shixian's wedding. Give you my motto: Keep healthy, but drink more!

To my parents, parents-in-law and my grandparents on both sides, you have dedicated so much and I would not be able to stand here without all your support. I wish you all good health. My dear grandfather, I am sure that you would be proud of me if you could witness this moment. May you rest in peace. To my beloved wife Danni, you are always there feeling me and encouraging me. Your love is like that mellow moonlight, making me tranquil, brave and determined. You make me a better man. I feel I was rushing through life before meeting you and I have been enjoying every moment of life since we were together. I decide not to say too much about you. Otherwise, I would get jealous by other men.



Static prediction



Dynamic prediction