



Universiteit  
Leiden  
The Netherlands

## Statistical methods for frailty models: studies on old-age mortality and recurrent events

Böhnstedt, M.

### Citation

Böhnstedt, M. (2021, November 30). *Statistical methods for frailty models: studies on old-age mortality and recurrent events*. Retrieved from <https://hdl.handle.net/1887/3243966>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3243966>

**Note:** To cite this publication please use the final published version (if applicable).

# Samenvatting

Frailty modellen blijven een belangrijke rol spelen in de analyse van survival data. Het introduceren van een positief random effect, de frailty, dat het risico om de betreffende gebeurtenis te ervaren beïnvloedt, kan ongeobserveerde heterogeniteit tussen individuen verklaren. In studies met multivariate survival data kunnen frailties ook gebruikt worden om de afhankelijkheid tussen observaties van hetzelfde individu of van verschillende eenheden van een cluster te verklaren. Een verscheidenheid van frailty modellen zijn voorgesteld in verschillende contexten, variërend van frailty proportional hazards modellen voor univariate survival data, tot shared frailty modellen voor geclusterde survival data, tot joint frailty modellen waarin twee of meer processen tegelijkertijd worden gemodelleerd. Hoewel er aanzienlijke vooruitgang is geboekt op dit gebied de afgelopen decennia, is verdere ontwikkeling van de methodologie vaak nodig om specifieke toepassingen aan te pakken.

In dit proefschrift breiden we de beschikbare statistische technieken voor schatten en toetsen in twee verschillende frailty modellen. In het eerste deel richten we ons op een frailty proportional hazards model om mortaliteit in volwassen mensen te bestuderen. Met het doel om methoden te identificeren die een goede beoordeling van het mortaliteitsproces op gevorderde leeftijd toestaan, bespreken we aspecten van parameterschatting, hypothesetoetsen, en onderzoeksopzet in dit model, evenals modelselectie. Het tweede deel betreft schatten en toetsen in een joint frailty model voor recurrenente gebeurtenissen en een terminale gebeurtenis in situaties waarin individuen pas later onder observatie komen (vertraagde binnenkomst), of waarin het recurrenente proces slechts met tussenpozen wordt geobserveerd. Nu volgt een gedetailleerder overzicht van de afzonderlijke hoofdstukken.

Hoofdstuk 1 geeft een korte introductie tot frailty modellering van overlevingsduur data in het algemeen, en biedt achtergrondinformatie over de modellen en de methoden die we bestuderen in de twee volgende delen van het proefschrift.

In het eerste deel van het proefschrift, dat bestaat uit hoofdstuk 2 tot en met 4, onderzoeken we verschillende aspecten van statistische studies over een demografisch fenomeen dat bekend staat als mortaliteitsvertraging. Deze vertraging van de sterftecijfers op gevorderde leeftijd wordt geanalyseerd in het kader van het gamma-Gompertz model. In dit frailty proportional hazards model ontstaat de vertraging als een effect van selectie, omdat de groep overlevenden tot op hogere leeftijd bestaat uit de sterkere individuen met een lager sterfterisico. De concurrerende situatie zonder sterftevertraging doet zich voor als de frailty variantie de waarde nul aanneemt. Het model wordt dan vereenvoudigd

tot het Gompertz model. In dit geval ligt de frailty variantie echter op de grens van zijn parameter ruimte. Deze niet-standaard situatie van een ‘parameter op de grens’ vereist aanpassingen van de statistische methoden die worden gebruikt voor het ontdekken van de potentiële vertraging van de sterftcijfers. Het uitvoeren van empirische onderzoeken naar mortaliteit op hoge leeftijd kan bovendien lastig zijn vanwege het gebrek aan data in de staart van de survival verdeling. Daarnaast kunnen steekproeven, aangezien een wetenschappelijke bevestiging van de leeftijd bij overlijden vaak verplicht is, maar van nature onderhevig is aan tijds- en kostenlimieten, worden begrensd om alleen overlevenden voorbij een zekere leeftijd te dekken. In het licht van deze uitdagingen bestuderen en vergelijken we verschillende benaderingen om mortaliteitsvertraging vast te stellen in het eerste deel van het proefschrift.

Hoofdstuk 2 onderzoekt de asymptotische eigenschappen van likelihood-schattingen in het gamma-Gompertz model, evenals het Akaike informatie criterium (AIC) voor dit model. Vanwege de grensparameter zijn de standaard resultaten van likelihood theorie niet geldig in dit geval. Onze afleidingen zijn gebaseerd op een kader van lokale alternatieven. Een bevinding van ons onderzoek is dat de maximum likelihood schatter in het gamma-Gompertz model asymptotisch een mengsel van twee verdelingen volgt als de onderliggende frailty variantie klein is. Evenzo heeft de toetsingsgrootheid van een likelihood ratio toets voor een frailty variantie van nul asymptotisch een mengsel van twee verdelingen. Een expliciete formule voor de geschatte lokale power van deze toets om een positieve frailty variantie te ontdekken wordt gepresenteerd. Bovendien bewijzen we dat de standaard formule voor de AIC niet van toepassing is op het gamma-Gompertz model, maar een onzuivere schatting geeft van de overeenkomstige Akaike informatie.

In hoofdstuk 3 bespreken we problemen van onderzoeksdesign voor onderzoeken naar mortaliteit op hoge leeftijd gebaseerd op het gamma-Gompertz model, en introduceren we informatiematen om dergelijke designs te evalueren. De design aspecten die we in overweging nemen omvatten de keuze van de leeftijdscategorie van een steekproef en de steekproefomvang. Op basis van de theorie van optimaal design definiëren we verschillende informatiematen als functies van de Fisher informatie matrix van het gamma-Gompertz model. Er worden interpretaties van deze maten gegeven, onder andere met betrekking tot de niet-standaard toestand van de grensparameter. We laten zien hoe de voorgestelde maten gebruikt kunnen worden om verschillende onderzoeksdesigns te beoordelen. De aanpak wordt aangevuld met berekeningen van het onderscheidend vermogen van een likelihood ratio toets om een vertraging in de sterftcijfers in specifieke design situaties te detecteren. Wij vinden dat de informatie die de data bevat en het onderscheidend vermogen van de toets, in de scenario's die wij bekijken, duidelijk verminderd zijn als alleen overlevenden tot toenemende hoge leeftijden in een studie worden opgenomen. In een toepassing van de voorgestelde methoden op sterftedata van Frans-Canadezen geboren eind 19e eeuw, onderzoeken we de mogelijke voordelen voor de statistische analyse als informatie over sterftes op de leeftijden 85-89 toegevoegd zou kunnen worden aan een bestaande dataset over overlevenden voorbij de leeftijd van 90.

In hoofdstuk 4 introduceren we “focused model selectie” als een nieuwe methode om sterftevertraging vast te stellen. Een focused informatie criterium (FIC) evalueert de model prestatie voor een specifieke parameter van belang, de focus parameter, en stelt ons daarom in staat om model prestatie op hoge leeftijden, waarvoor sterftevertraging het opvallendst is, op te lossen. We leiden formules van een FIC af om tussen twee kandidaat-modellen te kiezen die slechts met een grensparameter verschillen. In een simulatiestudie onderzoeken we hoe deze nieuwe versie van de FIC presteert bij het selecteren tussen het gamma-Gompertz model en het Gompertz model, afhankelijk van de keuze van de focus parameter. We zien dat, vergeleken met een aangepaste versie van de AIC die een bias correctie term bevat gebaseerd op de resultaten van hoofdstuk 2, de FIC een betrouwbaarder hulpmiddel is om een vertraging in de sterftecijfers te detecteren. Modelselectie met behulp van de nieuwe informatiecriteria wordt ook geïllustreerd met de Frans-Canadese sterftedata.

Het tweede deel van het proefschrift, dat bestaat uit hoofdstukken 5 en 6, gaat over de studie van recurrente gebeurtenissen in de aanwezigheid van een terminale gebeurtenis. Een joint frailty model wordt gebruikt om de potentiële afhankelijke censurering van het recurrente proces door de terminale gebeurtenis te verklaren. Het model bevat een aanvullende associatie parameter die bepaald of, en zo ja, in hoeverre hogere recurrence rates impliceren dat er een hoger of lager risico is om de terminale gebeurtenis te ondergaan. Tot nu toe werden schattingsmethoden in dit joint frailty model voornamelijk bestudeerd in ideale situaties, waarin de recurrentietijden precies bekend zijn en individuen geobserveerd worden vanaf het moment dat ze gevaar lopen om de betreffende gebeurtenissen mee te maken. Wij passen de statistische methoden aan aan twee andere veelvoorkomende situaties. In de eerste situatie zijn alleen interval tellingen van de recurrente gebeurtenissen, die de aantallen recurrenties in specifieke tijdsintervallen geven, beschikbaar. In de tweede situatie worden individuen pas in de studie opgenomen nadat de processen zijn begonnen, wat resulteert in truncatie van links. In beide situaties is parameter schatting gebaseerd op de marginale waarschijnlijkheid, die benaderd wordt met behulp van Gaussian quadrature. Voor de baseline rates nemen we stuksgewijs constante modellen aan.

In hoofdstuk 5 presenteren we de studies met interval tellingen van recurrente gebeurtenissen in een algemene situatie waarin de observatie-intervallen kunnen variëren tussen individuen. Naast de schattingsmethode passen we ook de scoretoets aan aan de situatie met interval tellingen, wat ons in staat stelt vóór het schatten van het joint model de associatie tussen de recurrente gebeurtenissen en de terminale gebeurtenis vast te stellen. Zowel de schattings- als de toetsingsprocedure presteren goed bij gesimuleerde data. We passen de methodes toe om te bepalen of asexuele reproductie en sterfte in het zeeorganisme *Eleutheria dichotoma* gerelateerd zijn.

In hoofdstuk 6 behandelen we het probleem van vertraagde binnenkomst in studies met recurrente gebeurtenissen met een terminale gebeurtenis. In zulke gevallen kunnen individuen alleen worden opgenomen in de studie als de terminale gebeurtenis voor hen nog niet heeft plaatsgevonden aan het begin van de studie. Als de recurrence rates en de

hazard van de terminale gebeurtenis geassocieerd zijn, kan deze selectie leiden tot verschillen tussen de frailty verdelingen in de steekproef en de onderliggende populatie. In simulatiestudies demonstreren we dat het negeren van de selectie effecten op de frailty verdeling in de constructie van de likelihood de schattingsresultaten kan vertekenen. Verder laten we zien hoe verschillende observatieschema's, en overeenkomstige patronen van incomplete informatie, de prestatie van de schattingsprocedure kunnen beïnvloeden. De voorgestelde methoden stellen ons in staat recurrenente urineweginfecties in een oudere bevolking te bestuderen met het gebruik van leeftijd als de belangrijkste tijdsschaal.