



Universiteit
Leiden
The Netherlands

Statistical modelling of repeated and multivariate survival data

Wintrebert, C.M.A.

Citation

Wintrebert, C. M. A. (2007, March 7). *Statistical modelling of repeated and multivariate survival data*. Department Medical Statistics and bio informatics, Faculty of Medicine / Leiden University Medical Center (LUMC), Leiden University. Retrieved from <https://hdl.handle.net/1887/11456>

Version: Corrected 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/1887/11456>

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

Samenvatting:

Statistisch Modelleren van Herhaalde en Multivariate Overlevingsgegevens

In dit proefschrift werden statistische modellen voor complexe overlevingsgegevens bestudeerd; zowel wanneer er *herhaalde* metingen als *verschillende* metingen per patiënt aanwezig waren. We gebruikten *frailty* modellen waarbij *random* effecten per patiënt geïntroduceerd werden om met de correlatie tussen de herhaalde en de *multivariate* metingen rekening te houden. Daarnaast beschouwden we gewogen en marginale overlevingsmodellen bij welke de correlatiestructuur ongedefinieerd bleef.

In hoofdstuk 2 ontwikkelden we tijdsafhankelijke frailty modellen en pasten die toe op overlevingsgegevens van patiënten uit het European Blood Marrow Transplant (EBMT) register, waarbij een met de tijd verminderend centrumeffect werd gevonden. Centrumeffecten worden gewoonlijk onderzocht met *constant shared frailty* modellen (Andersen et al., 1999), en dit veronderstelt dat dit effect constant aanwezig is tijdens follow-up, zelfs wanneer deze zeer lang duurt. Maar bij vele toepassingen is deze veronderstelling te beperkend. Daarom bestudeerden we methoden om de *constant frailty* modellen te generaliseren zodat de centrumeffecten variëren met de tijd. Deze modellen bleken realistischer. Wij vergeleken eveneens onze resultaten met de resultaten van het tijdsafhankelijke model van Paik et al. (1994). Hun model is moeilijk te berekenen en de parameters van hun model zijn moeilijk te schatten. Wij stelden twee eenvoudigere modellen voor en kregen bevredigende resultaten. Een duidelijk centrumeffect op de overleving na transplantatie werd gevonden en dit effect was vooral aanwezig in de eerste drie tot vier maanden na transplantatie, ook wanneer rekening werd gehouden met andere risicofactoren bij de analyse.

In hoofdstuk 3 analyseerden we overlevingsdata uit de ecologie. Wij gebruikten statistische methoden voor de analyse van overlevingsgegevens, waarbij met individuele variatie rekening gehouden wordt; dit om de correlatie tussen de overleving van (Kittiwake) meeuwen en hun herhaald gemeten broedpogingen vast te stellen. In onze analyse hielden we rekening met het feit dat de data rechts en interval gecensureerd waren. Een interessant punt van het model dat we ontwikkelden en toepasten, is dat we één enkele frailty in beide modellen voor de overleving en het broeden gebruikten, wat resulteerde in een spaarzamer model dan het model gebruikt door Cam and Monnat (2000). Verder werden onze modellen zowel geschat met een frequentiste me-

thode als met een Bayesiaanse methode met behulp van MCMC steekproeven in Winbugs (Spiegelhalter et al., 1996). Wij vonden een positieve correlatie tussen overleving en broeden bij vogels en dit resultaat is consistent met dat van Cam et al. (2002) die dezelfde data gebruikte maar andere modellen uit de ecologie.

In hoofdstuk 4 ontwikkelden we modellen om het verband tussen familieleden te analyseren met betrekking tot leeftijd bij aanvang van de ziekte van Huntington. Over het algemeen zijn genetische modellen voor overlevingsdata moeilijk te formuleren en de parameters van modellen moeilijk te schatten. Verscheidene statistische modellen werden ontwikkeld om de correlatie tussen familiair gerelateerde *failure type* gegevens te onderzoeken. Deze modellen, bijvoorbeeld dat van Li and Zhong (2002), zijn zeer complex, vooral voor grote en complexe stambomen, en de parameters van deze modellen zijn ook weer zeer moeilijk te schatten. Bovendien hangt het gevonden verband af van de gemaakte veronderstellingen. In hoofdstuk 4 gebruikten we een andere benadering. We toonden aan hoe martingaal residuen van een (marginaal) Cox model kunnen worden gebruikt om de aanwezigheid van een genetisch effect te schatten en om de genetische correlaties die afhangen van de genetische afstand in te schatten. Uit onze resultaten bleek dat een belangrijk verband bestaat tussen de leeftijden waarbij voor het eerst symptomen van de ziekte van Huntington optreden bij individuen van dezelfde familie, zelfs na correctie voor risicofactoren en voor de mutatie die oorzaak is voor de ziekte van Huntington. Dit verband is groter tussen broers en zussen en tussen ouders en kinderen, dan bij individuen met een grotere genetische afstand. Deze observatie suggereert het bestaan van één of meer genen, andere dan het gen van de ziekte van Huntington, die betrokken zijn bij het tijdstip waarop de ziekte van Huntington verschijnt. Onze nieuwe methode is eenvoudig en toepasbaar op stambomen van allerlei afmetingen, en is geïmplementeerd in het software programma S-plus.

In hoofdstuk 5 analyseerden we multivariate overlevingsgegevens van patiënten met arteriosclerose. *Multivariate frailty* modellen werden gebruikt om de correlatie te schatten tussen 3 processen die leiden tot cardiale, cerebrale en perifere arteriosclerose. Uitgangspunt was de veronderstelling dat de 3 processen verschillend waren. De correlaties tussen de 3 arteriosclerose processen werden geschat bij patiënten die twee of meer arteriosclerose gebeurtenissen hadden meegemaakt. Om deze correlaties te bepalen werden verscheidene *frailty* modellen ontworpen en toegepast. Moeilijkheden die zich bij de analyses in dit hoofdstuk voordeden, waren dat de patiënten geselecteerd waren vanwege het hebben meegemaakt van een eerste arteriosclerose gebeurtenis en dat een minderheid een tweede arteriosclerose gebeurtenis meemaakte tijdens de follow-up.

Onze resultaten vertonen een zeer sterke correlatie tussen herhaalde gebeurtenissen van dezelfde arteriosclerose soort. Klinisch gezien is het belangrijk dat onze resultaten suggereren dat het risico voor een gebeurtenis (eender welke) vooral afhangt van de leeftijd van de patiënt bij de eerste gebeurtenis, en niet van de soort van de gebeurtenis.

In hoofdstuk 6 schatten we marginale overlevingscurven in gebalanceerde en ongebalanceerde longitudinale studies. Verscheidene benaderingen, parametrische en non-parametrische, werden ontwikkeld om eventueel gecensureerde, herhaalde en ongebalanceerde data te analyseren. Twee klassen van modellen werden voorgesteld, namelijk gewogen en frailty modellen en de twee benaderingen werden vergeleken met betrekking tot het schatten van de marginale overlevingscurven.

De precisie van beide modellen voor ongebalanceerde data werd geevalueerd, alsmede de robuustheid van een slecht gespecificeerd model voor het random effect werd bestudeerd. De moeilijkheid lag hem daarin dat de keuze van de frailty verdeling ook de afhankelijkheid tussen herhaalde metingen bepaalde: wanneer de parameters van het model niet goed te schatten zijn of wanneer de specificaties van het model niet juist zijn, kan de marginale overlevingscurve moeilijk berekend worden. Het gewogen model was minstens zo goed als het frailty model. Het is onduidelijk waarom het gewogen model zulke goede resultaten geeft; misschien ligt het aan de eenvoud van het model of aan het gebrek van vooronderstellingen. In het algemeen bleek de marginale overlevingsfunctie moeilijk te bepalen te zijn, en omdat er weinig instrumenten beschikbaar zijn om de geldigheid te controleren, adviseren wij om verscheidene modellen naast elkaar te evalueren.

