Universiteit Leiden
The Netherlands

Statistical modelling of competing risks with incomplete data: with applications to allogeneic stem cell transplantation

Bonneville, E.F.

Citation
Bonneville, E. F. (2025, July 2). *Statistical modelling of competing risks with incomplete data: with applications to allogeneic stem cell transplantation*. Retrieved from https://hdl.handle.net/1887/4252266

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/1887/4252266

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

Nederlandse samenvatting

Vroeg of laat zullen ontbrekende data een uitdaging voor elke observationele medische studie vormen. Dit is vooral het geval bij het bestuderen van de tijd tot een bepaalde gebeurtenis, vanaf een klinisch relevant beginpunt. We kunnen bijvoorbeeld geïnteresseerd zijn in de tijd tot een infectie na een allogene hematopoietische stamceltransplantatie (alloSCT), een behandeling die voornamelijk wordt gegeven aan mensen met bloedkanker. Een studie moet uiteindelijk beëindigd worden, en op dat moment zullen sommige patiënten waarschijnlijk nog geen infectie hebben gehad. Omdat we niet weten of en wanneer ze geïnfecteerd zouden zijn als de onderzoeksperiode langer was geweest, is hun tijd tot infectie onbekend en worden ze aan het einde van de studie aangeduid als *rechtgecensureerd*. De tijd tot infectie zou ook onbekend zijn als een patiënt tijdens het onderzoek overlijdt voordat hij wordt geïnfecteerd. Overlijden wordt hier beschouwd als een *concurrerend risico*, omdat het een toekomstige infectie uitsluit.

Bovendien kan het interessant zijn om de relatie tussen patiëntspecifieke kenmerken en het optreden van een of meer concurrerende gebeurtenissen te bestuderen met behulp van statistische modellen. We kunnen bijvoorbeeld onderzoeken of de cumulatieve kans op infectie een jaar na een alloSCT kan worden voorspeld door de leeftijd van een patiënt bij alloSCT (een *baseline* covariaat), of door het aantal immuuncellen dat in de loop van de tijd in hun bloed circuleert (een *longitudinale* covariaat). Ontbrekende data in een of meer (soorten van) covariaten vormen een bedreiging voor de geldigheid van de modellen die worden gebruikt om de bovengenoemde vragen te beantwoorden. Naïeve uitsluiting van individuen met ontbrekende data kan de statistisch onderscheidingsvermogen drastisch verminderen en kan vertekende schattingen van doelgrootheden opleveren als de mogelijke oorzaken van de ontbrekende gegevens niet grondig worden overwogen. Adequate behandeling van ontbrekende waarden met behulp van principiële methoden zoals *multipele imputatie* (MI), in de context van regressiemodellen die rekening houden met concurrerende risico's, heeft weinig aandacht gekregen in de statistische literatuur. Dit proefschrift probeert deze onderzoekskloof te adresseren door het ontwikkelen, beoordelen en toepassen van statistische methodologie voor het omgaan met ontbrekende data in de context van alloSCT-studies met concurrerende risico's.

Hoofdstuk 2 geeft een hedendaags overzicht van hoe er momenteel wordt omgegaan met ontbrekende covariaatdata in de belangrijkste hematologische tijdschriften. Deze systematische review toonde aan dat ontbrekende covariaatdata veel voorkomen in klinische studies in de hematologie, maar dat bestaande richtlijnen met betrekking tot de behandeling en rapportage van ontbrekende data over het algemeen niet worden gevolgd. Bovendien werd MI, in tegenstelling tot eenvoudiger methoden zoals complete-case analyse en de missing indicator methode, zelden gebruikt.

In **Hoofdstuk 3** werden covariaat imputatiemodellen afgeleid die ongeveer 'compatibel' zijn met een of meer oorzaak-specifieke Cox proportional hazards analysemodellen. Een simulatiestudie toonde aan dat deze imputatiemodellen wat betreft het schatten van oorzaak-specifieke hazard ratio's over het algemeen slechter presteerden dan een eerder voorgestelde zogeheten substantive-model-compatible MI methode (SMC-FCS). Wat betreft het schatten van individu-specifieke cumulatieve incidentiefuncties presteerden beide methoden vergelijkbaar. De bovengenoemde ongeveer-compatibele MI methode werd toegepast in **Hoofdstuk 4** op een dataset van patiënten met myelofibrose die een alloSCT hebben ondergaan, waarbij het doel was om de invloed van gedeeltelijk ontbrekend body mass index en comorbiditeiten op het oorzaak-specifieke risico van overlijden voor ziekteterugval te beoordelen. Dit hoofdstuk vormt ook een casestudy met betrekking tot de imputatie van zogenaamde 'afgeleide' covariaten (d.w.z. de combinatie van twee of meer direct gemeten variabelen).

In **Hoofdstuk 5** werd nieuwe SMC-FCS methodologie ontwikkeld die het mogelijk maakt om ontbrekende covariaten compatibel te imputeren met een Fine–Gray analysemodel, zonder een model te hoeven specificeren voor de concurrerende gebeurtenis(sen). Deze methode werd beoordeeld in een simulatiestudie, waarin er scenario's waren waarbij de imputatieprocedure compatibel was met het onjuiste data-genererende mechanisme (d.w.z. waar proportionaliteit gold op de oorzaak-specifieke hazard schaal in plaats van op de subdistributie hazard schaal). Hier bleek censurering een belangrijke rol te spelen, door de impact van analysemodel-mispecificatie in de imputatiefase te verzachten.

Hoofdstuk 6 geeft een beknopt overzicht van data-genererende mechanismen waarbij een Fine–Gray model correct is voor ten minste één concurrerende gebeurtenis. De belangrijkste conclusie van dit hoofdstuk was dat men de voorkeur moet geven aan oorzaak-specifieke hazard modellen boven meerdere Fine–Gray modellen wanneer meer dan één concurrerende gebeurtenis van belang is, omdat er geen data-genererend mechanisme is waarvoor de aanname van proportionele subdistributie hazards tegelijkertijd geldt voor alle gebeurtenissen (tenzij aanvullende aannames worden gemaakt). Dit hoofdstuk biedt ook inzichten met betrekking tot SMC-FCS met een Fine–Gray analysemodel, waarbij een model voor de concurrerende gebeurtenis alleen voor imputatiedoeleinden zou kunnen worden gespecificeerd.

Ten slotte toont **Hoofdstuk 7** het gebruik van zogenoemde *joint models* voor het analyseren van de trajecten van het aantal immuuncellen voor een cohort acute

leukemiepatiënten in de eerste 6 maanden na een alloSCT met T-celdepletie. De resultaten onderstreepten het belang van het in aanmerking nemen van concurrerende risico's in het tijd-tot-gebeurtenis submodel, waarbij met name de 'huidige waarde' van het aantal CD4+ cellen tegengestelde effecten heeft op de oorzaak-specifieke risico's van graft-versus-hostziekte (GvHD) en ziekte terugval.