

# Valkuilen en oplossingen bij de overlevingsduuranalyse in hematologische studies

Pitfalls and solutions in survival analysis for hematological studies

dr. L.C. de Wreede<sup>1</sup> en prof. dr. H. Putter<sup>2</sup>

## SAMENVATTING

De belangrijkste uitkomstmaat voor klinische hematologische studies is overlevingsduur. Het deelgebied van de statistiek dat deze uitkomsten analyseert, is de overlevingsduuranalyse ('survival analysis'). Ook andere uitkomsten waarbij niet alleen van belang is of ze optreden, maar ook wanneer, kunnen met de methoden uit dit domein worden geanalyseerd. De analyse van deze uitkomsten brengt speciale moeilijkheden met zich mee, omdat het tijd kost om gebeurtenissen die verspreid over de tijd optreden waar te nemen. De belangrijkste voorwaarde voor het berekenen van een overlevingskromme zonder 'bias' – niet-informatieve censurering – wordt besproken, alsmede een oplossing voor de situatie waarin verschillende eindpunten, die niet tegelijkertijd kunnen voorkomen, van belang zijn. Ook worden verschillende methoden voor het vergelijken van de uitkomsten van twee groepen uitgelegd, waarbij ook de situatie wordt besproken dat aan het begin van de studie nog niet bekend is wie in welke groep terecht zal komen.

(NED TIJDSCHR HEMATOL 2017;14:118-24)

## SUMMARY

The most important outcome of clinical hematological studies is survival. The statistical field that deals with such outcomes is survival analysis duration. Also other outcomes where not only the occurrence in itself, but also the timing of the events is important, can be analyzed with methods developed in this field. The analysis of time-to-event outcomes brings special challenges because of the fact that 'it takes time to observe time'. The most important condition for unbiased estimation of a survival curve – non-informative censoring – is discussed, as well as methods for competing risks, the situation where multiple, mutually exclusive, endpoints are of interest. Methods for comparison of survival outcomes between groups are reviewed, including the case where it is not yet known at baseline to which group a patient will belong later during follow-up (time-dependent covariates).

## INLEIDING

De belangrijkste uitkomstmaat voor klinische hematologische studies is overlevingsduur. De uiteindelijke effectiviteit van een behandeling, of relevantie van een risicofactor, wordt bepaald door de invloed op de overleving van patiënten. Het is zeker dat alle patiënten in

de toekomst zullen overlijden, maar het is van tevoren niet bekend wanneer dit zal gebeuren. Het deelgebied van de statistiek dat deze uitkomsten analyseert, is de overlevingsduuranalyse ('survival analysis'). Ook andere uitkomsten waarbij niet alleen van belang is of ze optreden, maar ook wanneer, zoals remissie, recidief,

<sup>1</sup>universitair docent/assistant professor, <sup>2</sup>hoogleraar/full professor, afdeling Medische Statistiek en Bioinformatica, Leids Universitair Medisch Centrum. Correspondentie graag richten aan mw. dr. L.C. de Wreede, universitair docent, afdeling Medische Statistiek en Bio-informatica, Leids Universitair Medisch Centrum, Einthovenweg 20, 2333 ZC Leiden, tel: 071 526 97 10, e-mailadres: l.c.de\_wreede@lumc.nl

Belangenconflict: geen gemeld. Financiële ondersteuning: geen gemeld.

**Trefwoorden:** 'competing risks', overlevingsduuranalyse, statistiek

**Keywords:** competing risks, statistics, survival analysis

complicaties en tweedelijnsbehandeling, kunnen met de methoden uit dit domein worden geanalyseerd.<sup>1,2</sup> Dergelijke uitkomsten worden altijd uitgedrukt als tijdsafhankelijke kansen, bijvoorbeeld de kans op overleving ('overall survival') op 1 jaar na het begin van de behandeling was 85%, of de kans op in leven zijn zonder nieuw recidief ('relapse-free survival') was 70% op 2 jaar na het bereiken van complete remissie. Zonder die tijdstoevoeging zijn uitspraken over percentages betekenisloos: op het starttijdstip is de kans op een goede uitkomst (dus in leven zijn of in leven zijn zonder recidief in de twee voorbeelden hierboven) per definitie 100%. Op lange termijn is de kans op in leven zijn voor elke groep patiënten (of welke levende wezens ook) 0%.

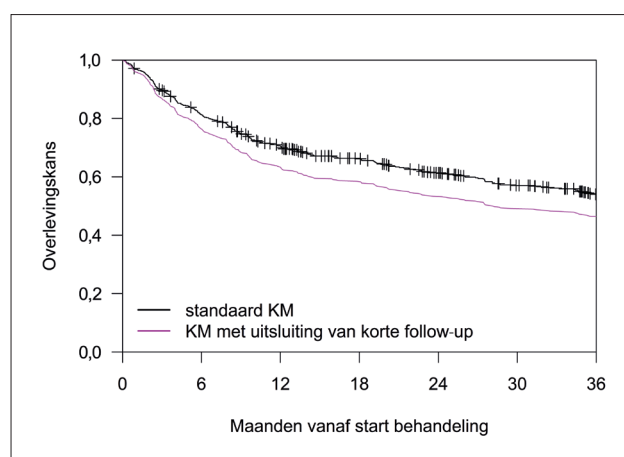
## DE DATASET

Alle valkuilen en oplossingen in dit artikel worden geïllustreerd aan de hand van een dataset die is afgeleid van een echte dataset, maar enigszins veranderd en geabstraheerd om de algemene principes uit te leggen. De dataset bevat informatie over 600 patiënten met een vorm van leukemie die ofwel de standaardbehandeling ofwel een experimentele behandeling hebben gekregen. Ze worden gevolgd vanaf het begin van hun behandeling. Twee vormen van mislukken van de behandeling worden bekeken: overlijden en recidief/progressie van de ziekte.

## CENSURERING

De grootste moeilijkheid bij het schatten van overlevingsduurkansen is dat het tijd kost om tijd waar te nemen. Bijna altijd is het onmogelijk om lang genoeg te wachten totdat bij alle patiënten in een studie de relevante uitkomst is opgetreden. Bovendien zijn er vaak patiënten die niet tot het eind van de studie zijn gevolgd ('loss to follow up'), bijvoorbeeld als gevolg van een verhuizing. Van die patiënten weten we wel dat ze gedurende een bepaalde tijd kans hebben gehad op de uitkomst en dat die toen niet is opgetreden, maar niet wat er daarna met hen is gebeurd. Dit verschijnsel heet censurering. In het algemeen is die periode ook nog voor alle patiënten in de dataset verschillend, omdat patiënten elk op hun eigen moment met de behandeling beginnen en hun follow-up beëindigen.

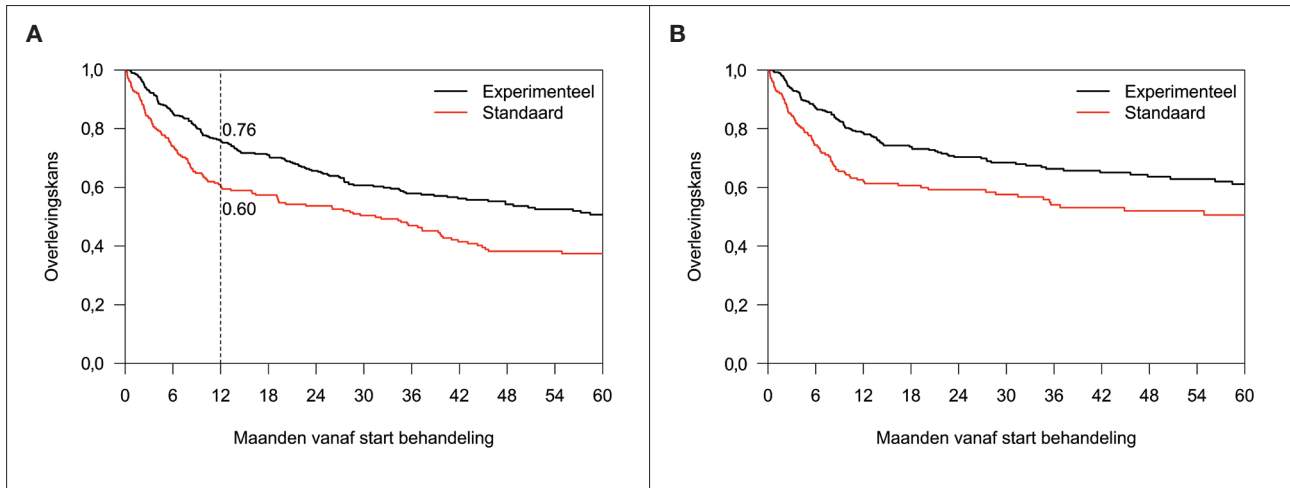
Kunnen deze patiënten eenvoudigweg worden genegeerd in de analyse omdat we niet weten wanneer ze zijn overleden? Nee, dat is niet mogelijk, omdat dat een 'bias' in de analyse zou geven: een systematisch verschil tussen de echte uitkomst en de uitkomst die we schatten in de analyse. Dat komt doordat we dan wel de patiënten meenemen die relatief snel zijn overleden, maar niet



**FIGUUR 1.** Standaard Kaplan-Meier-kromme (zwart, plus-tokens geven gecensureerde waarnemingen aan) en de Kaplan-Meier-schatter met uitsluiting van patiënten met follow-up minder dan 36 maanden die aan het eind van hun follow-up nog in leven waren (roze).

degenen die ongeveer even lang zijn gevolgd, maar na die tijd nog in leven waren. Dat leidt dus tot een overschatting van de kans op overlijden. *Figuur 1* illustreert dit verschijnsel. De zwarte kromme geeft de correcte schatting van de overlevingsduur, uitgerekend met behulp van de Kaplan-Meier-methode, waarbij alle gecensureerde patiënten met een verticaal streepje zijn aangeduid.<sup>3</sup> De roze kromme illustreert de foutief geschatte kansen nadat alle patiënten die binnen drie jaar gecensureerd zijn uit de dataset zijn verwijderd. Het verschil tussen de twee krommes is de 'bias'.

Het schatten van uitkomsten in de overlevingsduur-analyse wordt dus gecompliceerd doordat alle informatie van alle patiënten, zowel de tijd dat ze gevolgd zijn en of ze al dan niet de uitkomst hebben ondervonden, moet worden geïntegreerd. Dit gebeurt in de drie belangrijkste methoden op dit gebied: de Kaplan-Meier-methode voor het schatten van kansen, de log-ranktoets voor het vergelijken van overlevingskrommes en het Cox-model voor proportionele hazards voor het schatten van de impact van verschillende risicofactoren op de uitkomsten.<sup>4</sup> Hierbij is één aanname essentieel: de patiënten die later in de tijd nog worden gevolgd, representeren de patiënten die een kortere follow-up hebben. Met andere woorden, als de follow-up van een patiënt op bijvoorbeeld 1 jaar na het begin van de behandeling wordt beëindigd, dan zijn de kansen op overlijden na dat jaar voor deze patiënt gelijk aan die van een patiënt die nog wel wordt gevolgd. Dit is een redelijke aanname als de reden van censurering onaf-



**FIGUUR 2. A.** De standaard Kaplan-Meier-krommes (voor experimentele en standaardbehandeling). In de figuur zijn de overlevingskansen op 1 jaar na start van de behandeling aangeduid. **B.** De foutieve Kaplan-Meier-krommes waar patiënten met recidief gecensureerd zijn op het moment van recidief.

hankelijk is van de prognose van de patiënt, bijvoorbeeld omdat de studie wordt beëindigd. Als bijvoorbeeld patiënten met een goede prognose minder goede follow-up hebben, omdat ze minder vaak ter controle naar een arts gaan, gaat die aanname niet op en geeft de Kaplan-Meier-methode geen geldige schattingen. Ook andere methoden kunnen dit probleem niet oplossen, want er zijn nu eenmaal geen gegevens van patiënten na hun moment van censurering. Om deze reden is het van groot belang 'loss to follow up' in een studie zo veel mogelijk te beperken en vooral mechanismen die censurering voor sommige soorten patiënten aannemelijker maken dan voor andere tegen te gaan.

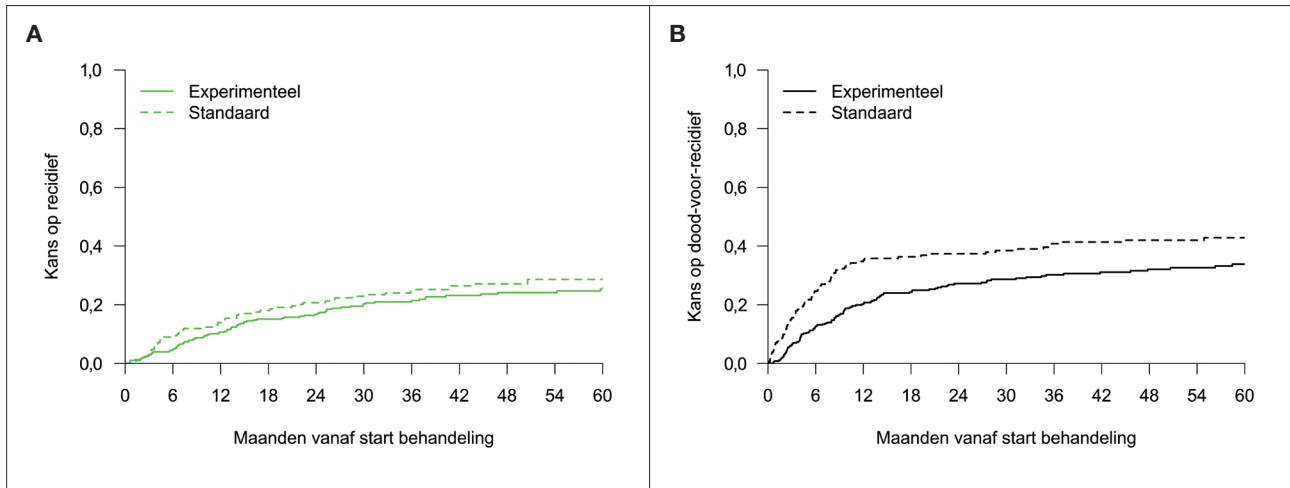
### INFORMATIEVE CENSURERING

In de praktijk is het niet altijd mogelijk 'loss to follow up' te voorkomen, zeker niet in studies gebaseerd op retrospectieve dataverzameling. Soms kiezen onderzoekers er echter zelf voor om gegevens te censureren, wat kwalijke gevolgen kan hebben voor de analyse. Ze willen bijvoorbeeld weten wat de effectiviteit van de standaard- en de experimentele behandeling is als eerstelijnsbehandelingen. Na een recidief ontvangen veel van deze leukemiepatiënten tweedelijnsbehandelingen, zoals allogene stamceltransplantatie, die extra kansen op overlijden met zich meebrengen. Daarin zijn de onderzoekers niet geïnteresseerd en daarom redeneren ze dat ze patiënten moeten censureren op het moment van hun eerste recidief. Dat lijkt in eerste instantie misschien een goed idee, maar hiermee wordt niet meer voldaan aan de bovengenoemde aanname van niet-

informatieve censurering. De uitkomsten van de patiënten met een slechtere prognose – die namelijk het gevolg is van hun recidief – worden systematisch uit de analyse gehaald. Bovendien hebben de resulterende krommen geen betekenis: die representeren de overlevingskansen in een wereld waar leukemiepatiënten geen recidief kunnen krijgen – overduidelijk een aantrekkelijk, maar onrealistisch perspectief. De vergelijking tussen de twee behandelingen wordt nog verder bemoeilijkt als de kansen op recidief in de twee groepen verschillend zijn. *Figuur 2* illustreert de impact van het censureren van patiënten bij recidief: A laat de goede krommes zien, B de verkeerde, te optimistische krommes. Hieruit volgt een eenduidig advies: censureer patiënten niet op grond van een tussentijdse gebeurtenis, zoals een recidief of remissie, omdat dat bijna altijd leidt tot informatieve censurering. Bij de evaluatie van de impact van een behandeling op sterfte horen dan ook alle overlijdensgevallen te worden meegenomen, ook die na tweedelijnsbehandeling, omdat die indirect voortvloeien uit de eerstelijnsbehandeling.

### METHODEN VOOR 'COMPETING RISKS'

Is het dan niet mogelijk de sterfte na de eerstelijnsbehandeling te scheiden van die na de tweedelijnsbehandeling? Ja, dat is wel mogelijk, als we bovenbeschreven methoden zo uitbreiden dat we niet naar één uitkomst kijken (overlijden, of recidief of dood (wat het eerste komt) in het geval van recidiefvrij overleven), maar naar verschillende, bijvoorbeeld recidief en dood voor recidief als twee aparte uitkomsten. Dit kan met



**FIGUUR 3.** Cumulatieve-incidentiekrommes voor recidief **(A)** en voor dood voor recidief **(B)**, voor experimentele en standaardbehandeling. De experimentele behandeling lijkt vooral de kans op dood voor recidief te verminderen.

methoden voor concurrerende risico's ('competing risks'): ze concurreren met elkaar omdat het optreden van de één het per definitie onmogelijk maakt dat de ander nog gaat plaatsvinden. Om bijvoorbeeld te zien wat de impact is van twee behandelingen op ziekteactiviteit, kunnen we de cumulatieve incidentie van recidief berekenen, dat wil zeggen de tijdsafhankelijke kans op een recidief. Patiënten die overlijden zonder dat ze eerst een recidief hebben gekregen, bijvoorbeeld door toxiciteit van de behandeling, worden dan niet gecensureerd, maar meegenomen in de berekening van de cumulatieve incidentie van dood voor recidief ('non-relapse mortality'). Op deze manier worden de twee componenten van het einde van recidiefvrij overleven – namelijk recidief en dood voor recidief – van elkaar gescheiden.

Een belangrijk voordeel daarvan is dat het ook mogelijk wordt de invloed van behandelingen op ieder van die twee componenten te analyseren. Het is heel goed mogelijk dat de experimentele behandeling geen netto-effect heeft op recidiefvrij overleven, maar wel degelijk anders werkt dan de standaardbehandeling: grotere effectiviteit (dus lagere kans op een recidief) kan goed samengaan met grotere toxiciteit (hogere kans op dood voor recidief). *Figuur 3* toont een situatie waar de winst van het nieuwe middel vooral zit in verminderde sterfte na eerstelijnsbehandeling.

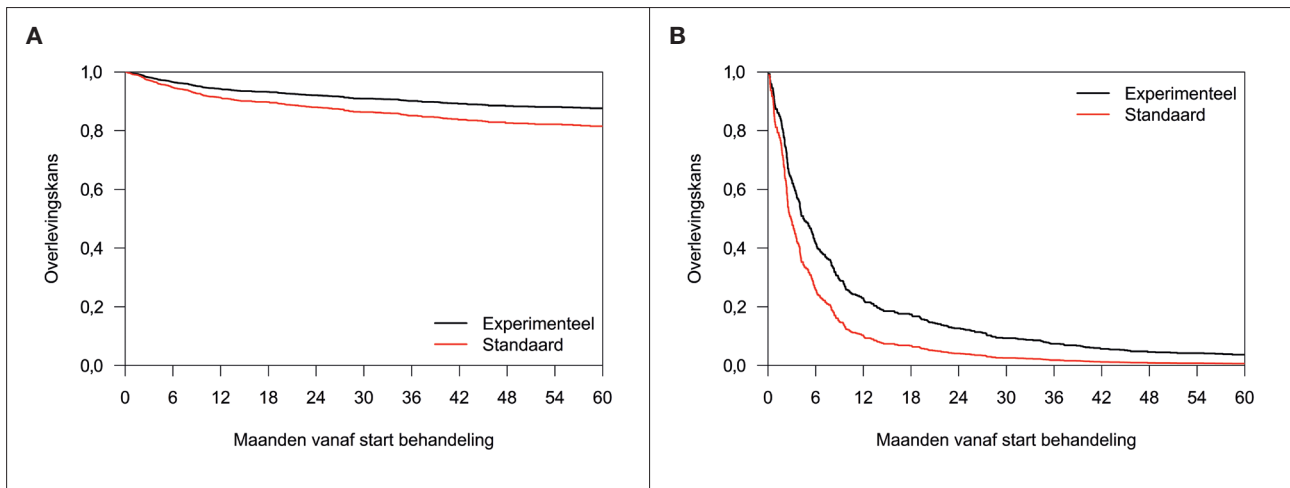
### VERGELIJKING VAN BEHANDELINGEN

*Figuur 2A* laat de geschatte overlevingskrommes zien voor de standaardbehandeling en de experimentele behandeling. De geschatte percentages voor respectieve-

lijk de experimentele en de standaardbehandeling zijn op 1 jaar 76,1% en 60,4%, en op 3 jaar 57,9% en 46,9%. Kunnen we hieruit concluderen dat de overleving bij de experimentele behandeling beter is dan bij standaardbehandeling? Ten eerste moeten we systematische verschillen van toevallige verschillen in de steekproef zien te onderscheiden. Dit doen we met een statistische toets. Ten tweede bestaan de overlevingskrommes uit geschatte overlevingskansen voor beide groepen op een heleboel tijdstippen. Dat zou potentieel heel veel toetsen opleveren, terwijl we in principe geïnteresseerd zijn in de vergelijking van de hele overlevingskrommes van de behandelingen. De toets die wordt gebruikt om overlevingskrommes met elkaar te vergelijken heet de log-ranktoets. Bij de overlevingskrommes van *Figuur 2A* is de p-waarde van de log-ranktoets 0,0002, een significant verschil dus tussen de twee overlevingskrommes.

### PROPORTIONELE-HAZARDSMODELLEN

De log-ranktoets geeft alleen een statistische uitspraak over de vraag of de overleving gelijk is voor twee behandelingen. De volgende vraag is: hoe groot is het effect van een behandeling? Een manier om deze vraag te beantwoorden is om terug te gaan naar de geschatte overlevingskansen op 3 jaar. Het verschil tussen de 46,9% van de standaard- en de 57,9% van de experimentele behandeling, oftewel 11%, kan worden gezien als een schatting van het effect van een behandeling op overleving na 3 jaar. Het voordeel van deze manier om het effect van de behandeling te kwantificeren is dat men eenvoudig het aantal patiënten kan berekenen dat men



**FIGUUR 4.** Overlevingskrommes voor twee behandelingen met een hazard ratio, net als in *Figuur 2*, van 1,55. Ten opzichte van *Figuur 2* is de baseline hazard met een factor 5 verlaagd (**A**) en verhoogd (**B**).

zou moeten behandelen met de experimentele behandeling (ten opzichte van standaard) om een overlijden binnen 3 jaar te voorkomen, in het Engels: ‘number needed to treat’. In dit geval komt dat uit op 9 (namelijk  $100/11$ ). Het nadeel van deze benadering is dat de keuze van 3 jaar tamelijk arbitrair is; na 1 jaar is het verschil tussen de overlevingskansen 15,7% en na 5 jaar 12,1%. Het is handiger om één enkele effectmaat te hebben die geldt voor de hele tijdsperiode.

Een populaire manier om het effect van een behandeling op overleving te kwantificeren is via de hazard ratio. Om te begrijpen wat een hazard ratio is, moeten we eerst ingaan op wat een ‘hazard rate’ is. De ‘hazard rate’, vaak afgekort tot hazard, beschrijft het risico op overlijden in de nabije toekomst (voor mensen die op dat moment nog in leven zijn). De ‘hazard rate’ verandert met de tijd; over het algemeen zal bijvoorbeeld de ‘hazard rate’ op de lange termijn door veroudering toenemen: de gemiddelde tachtigjarige heeft een grotere kans de komende drie maanden te overlijden dan de gemiddelde veertigjarige. Daarnaast kan een agressieve behandeling zoals allogene stamceltransplantatie ervoor zorgen dat de ‘hazard rate’ in de initiële periode na de behandeling verhoogd is. De hazards en de overlevingskansen zijn via wiskundige formules in elkaar uit te drukken, dat wil zeggen dat ze in feite dezelfde informatie bevatten vanuit verschillende perspectieven.

Bekijken we twee behandelingen, dan zullen de ‘hazard rates’ van beide behandelingen met de tijd variëren. Bij het proportionele-hazardsmodel wordt een belangrijke aanname gedaan, namelijk dat de hazard ratio, dat wil zeggen de verhouding tussen de twee ‘hazard rates’,

constant blijft. Het is een model, dat wil zeggen dat er een bepaalde structuur aan de data wordt opgelegd. In het voorbeeld van de vergelijking van de twee behandelingen in *Figuur 2A* is de hazard ratio van de standaardten opzichte van experimentele behandeling gelijk aan 1,55, met een 95%-betrouwbaarheidsinterval van 1,23 tot 1,94. Dit betekent dat op elk moment in de tijd het directe risico om dood te gaan voor de patiënten die de standaardbehandeling ondergaan 1,55 keer zo hoog is als voor degenen die de experimentele behandeling krijgen. Bij het vermelden van een hazard ratio moet het duidelijk zijn wat met wat wordt vergeleken. Een hazard ratio van 1,55 zegt niets als niet wordt vermeld wat de referentiegroep is en welke andere groep daarmee wordt vergeleken. De hazard ratio van de experimentele behandeling ten opzichte van standaard is 0,65 ( $1/1,55$ ). De hazard ratio is een vermenigvuldigingsfactor van de hazard en niet van de overlevingskansen. Een hazard ratio van 1,55 wil dus *niet* zeggen dat bij de standaardbehandeling de kans om binnen 3 jaar te overlijden 1,55 keer zo hoog is als bij de groep met de experimentele behandeling. Een hazard ratio van 1,55 kan zich namelijk op verschillende manieren manifesteren in twee overlevingskrommes. Vast staat dat de overlevingskromme van de experimentele behandeling boven die van de standaardbehandeling moet liggen, maar hoe ver de krommes van elkaar afdiggen, hangt af van de ‘baseline hazard’, het risico in de referentiegroep. Als de ‘baseline hazard’ hoog is, dan is de overleving van de referentiegroep laag en de overleving van de andere groep nog lager; als de ‘baseline hazard’ laag is, dan is de overleving van de referentiegroep hoog (dicht bij 1)

en zal de overleving van de andere groep slechter zijn, maar mogelijksterwijs nog steeds dicht bij 1. *Figuur 4* laat twee keer een vergelijking van twee overlevingskrommes zien, beide met een hazard ratio van 1,55, als in *Figuur 2A*. De 'baseline hazard' van *Figuur 2A* is daarbij eerst vijf keer zo hoog gemaakt (*Figuur 4A*) en daarna vijf keer zo laag (*Figuur 4B*). De krommes zien er heel verschillend uit, maar in alle drie de gevallen is de hazard ratio 1,55.

### NIET-PROPORTIONELE HAZARDS

De cruciale aanname van het proportionele-hazardsmodel is dat de ratio van twee hazardkrommes (die beide met de tijd variëren) constant is. Dit betekent dat wordt aangenomen dat het effect van een behandeling vlak na de start hetzelfde is als later in de follow-up. Deze aanname blijkt in de praktijk vaak redelijk te werken, vooral als de bestudeerde tijdsduur niet te lang is. In een aantal situaties is de proportionele-hazardsaanname echter niet realistisch. Een belangrijk voorbeeld is het vergelijken van een standaardbehandeling met een meer ingrijpende behandeling (zoals allogene stamceltransplantatie), die in het begin wellicht nadelig is, maar op langere termijn meer perspectief op genezing biedt. De hazard ratio van experimenteel ten opzichte van standaard zal in dit voorbeeld in het begin hoger dan 1 zijn (door de hogere posttransplantatiemortaliteit) en later lager dan 1 (als recidieven en daaruit voortkomende doden worden voorkomen). Als in een dergelijke situatie een proportionele-hazardsmodel wordt gebruikt, zal de geschatte hazard ratio (onder de incorrecte proportionele-hazardsaanname) een gemiddelde van de hazard ratio's over de tijd zijn. De uitkomst van de analyse zal dan afhangen van de duur van de follow-up; bij kortere follow-up zal de experimentele behandeling nadeliger uit de bus komen dan bij langere follow-up. Het is dus van belang de proportionele-hazardsaanname te controleren en het eindtijdstip van de analyse te specificeren.

### TIJDSAFHANKELIJKE COVARIATEN

Tot dusver hebben we twee behandelingen vergeleken in een proportionele-hazardsmodel. Behandeling fungeert dan als zogenoemde covariaat in het regressiemodel. Vaak zijn onderzoekers geïnteresseerd in het effect van een covariaat waarvan de waarde zelf met de tijd kan veranderen, een zogenoemde tijdsafhankelijke covariaat. Voorzichtigheid is geboden bij de analyse van de impact van dergelijke tijdsafhankelijke covariaten. Als in de analyse geen rekening wordt gehouden met het feit

dat de waarde van de tijdsafhankelijke covariaat niet constant is, dan kunnen volstrekt verkeerde conclusies worden getrokken. Als voorbeeld kijken we naar het effect van recidief op overleving. Een voor de hand liggende manier om het effect van recidief op overleving te schatten is met een proportionele-hazardsmodel met recidief als gewone covariaat. We maken dan feitelijk twee groepen: één groep met recidief en één zonder. Deze vergelijking gaat echter mank, omdat de patiënten in de recidiefgroep alleen in deze groep terecht kunnen komen als ze niet zijn overleden voor een recidief onderzonden te hebben. De mensen in deze groep zijn in zekere zin onsterfelijk tot het moment van hun recidief. De zo ontstane 'bias' staat bekend als 'immortal time bias'. Als we toch deze analyse doen, zouden we een hazard ratio van 1,08 krijgen voor de recidiefgroep ten opzichte van de groep zonder recidief, met een 95%-betrouwbaarheidsinterval van 0,84 tot 1,38 ( $p=0,58$ ). De (onterechte) conclusie zou zijn dat een recidief niet of nauwelijks invloed heeft op sterfte.

Wat kunnen we dan doen om een correcte vergelijking te verkrijgen? De eerste mogelijkheid is een zogenoemd proportionele-hazardsmodel met tijdsafhankelijke covariaten. Dit model gebruikt op elk tijdstip de recidief-status op dat moment voor de vergelijking van de hazards. Toepassing van dit model geeft een hazard ratio van 4,53 voor recidief ten opzichte van geen recidief. Dit betekent dat als we op elk moment in de tijd twee (levende) patiënten vergelijken van wie de een al wel en de ander (nog) geen recidief heeft gehad, de patiënt met recidief een 4,53 keer zo hoog direct risico heeft om dood te gaan als de patiënt zonder recidief. De tweede mogelijkheid is 'landmarking'. Bij deze methode wordt een landmarktijdstip bepaald. Op dat landmarktijdstip worden de patiënten die op dat moment nog in leven zijn beschouwd en wordt een vergelijking gemaakt tussen een groep die op dat moment wel en een andere groep die dan nog geen recidief heeft gehad; patiënten die dan al overleden zijn of een korte follow-up hebben, leveren dus geen bijdrage aan deze analyse. Met een landmark op 1 jaar resulteert dit in een hazard ratio van 3,93, met een 95%-betrouwbaarheidsinterval van 2,56 tot 6,03. Beide methoden zijn correct, maar bekijken de vergelijking vanuit een verschillend perspectief. Het proportionele-hazardsmodel geeft een samenvatting van de invloed van recidief op uitkomst over de hele tijdsperiode van de follow-up, het landmarkmodel concentreert zich op de invloed van recidieven die zijn opgetreden voor het landmarktijdstip op de latere gebeurtenissen.

## AANWIJZINGEN VOOR DE PRAKTIJK

- 1 Uitkomsten waarbij het tijdstip van optreden van belang is, vereisen speciale statistische technieken.
- 2 Censurering van patiënten mag niet samenhangen met hun prognose.
- 3 Als er meer dan één eindpunt kan optreden, zijn ‘competing risks’-modellen nodig.
- 4 Groepen patiënten kunnen niet worden ingedeeld op grond van informatie uit de toekomst.

## PRAKTISCHE TIPS

- 1 Een goed startpunt voor het ontwikkelen van meer inzicht in de statistiek zoals benodigd voor hematologische studies is te vinden in de *EBMT Statistical Guidelines* van Simona Iacobelli.<sup>2</sup>
- 2 De meeste hier beschreven analyses kunnen worden uitgevoerd in SPSS. ‘Competing risks’-analyses kunnen bijvoorbeeld worden uitgevoerd in het gratis softwarepakket R.

## CONCLUSIE

Uitkomsten waarin zowel het optreden van een gebeurtenis als het tijdstip daarvan worden geanalyseerd, vereisen speciale statistische technieken. Zowel bij het zelf verrichten van analyses als bij de interpretatie van resultaten in artikelen is het van belang bewust te zijn van aannames die aan de meest gebruikte technieken ten grondslag liggen: die van de niet-informatieve censurering en die van proportionaliteit van de hazards. Als die aannames niet waar zijn, kan dat grote gevolgen hebben voor de interpretatie van de uitkomsten. Een ander belangrijk punt betreft de analyse van samengestelde uitkomsten zoals recidiefvrij overleven. Overlevingskansen voor samengestelde uitkomsten zijn gemakkelijker te berekenen dan aparte uitkomsten zoals die in een ‘competing risks’-model worden geanalyseerd. Om de mechanismen die achter falen liggen te begrijpen, is het echter belangrijk om verschillende oorzaken van falen en de invloed van risicofactoren daarop van elkaar te onderscheiden.

Wij bevelen voor proportionele-hazardsmodellen aan om niet alleen naar de hazard ratio te kijken, maar ook

naar de ‘baseline hazard’, omdat alleen de combinatie van die twee een goede maat van het effect op de uitkomst van een behandeling of risicofactor biedt.

## DANKWOORD

De auteurs bedanken Michel van Gelder (MUMC) en Johannes Schetelig (Universiteitsziekenhuis Technische Universiteit Dresden/DKMS Clinical Trials Unit, Dresden, Duitsland) voor hun bijdragen aan dit artikel.

## REFERENTIES

1. Klein JP, Moeschberger ML. Techniques for censored and truncated data. New York: Springer-Verlag, 2003.
2. Iacobelli S, on behalf of the EBMT Statistical Committee. Suggestions on the use of statistical methodologies in studies of the European Group for Blood and Marrow Transplantation. *Bone Marrow Transplant* 2013;48:S1-37.
3. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Statistical Assoc* 1958;53:457-81.
4. Cox DR. Regression models and life-tables. *J Royal Statistical Soc Series B (Methodological)* 1972;34:187-220.

ONTVANGEN 24 OKTOBER 2016, GEACCEPTEERD 13 FEBRUARI 2017.

ALLE GEPUBLICEERDE ARTIKELLEN KUNT U VINDEN OP ONZE WEBSITE:

**WWW.ARIEZ.NL**

Tevens kunt u daar zoeken naar artikelen die in onze andere tijdschriften zijn gepubliceerd.