

Een protocol ter harmonisering van het gebruik van cohort data in survival studies

Ronald B. Geskus
GG & GD, Amsterdam

April 1998

De meeste survival studies naar HIV infectie en AIDS kijken naar AIDS, dood door AIDS of dood (algemeen) als eindpunt. Tot nu toe beslist iedereen zelf in zijn/haar analyse of bepaalde “moeilijke gevallen” rechts gecensureerd worden dan wel een event (AIDS/dood) krijgen. De hierbij gehanteerde criteria worden niet of slechts ten dele expliciet gemaakt. Daarnaast is er de overweging welke personen in de analyse meegenomen worden. Het onderstaande protocol maakt dergelijke keuzes op basis van een aantal duidelijke criteria, en zal daardoor meer harmonie in de gebruikte data brengen en tijdsbesparend werken. De gehanteerde criteria gelden in eerste instantie voor survival analyses met seroconversie als nulpunt; de criteria voor analyses van de predictieve waarde van immunologische, virologische en klinische markers worden apart genoemd. Voor klinische trials of case-control studies kunnen weer andere criteria gelden. Hoewel volledig toegespitst op de data uit de cohort studie onder homoseksuele mannen, zullen de meeste overwegingen in het tweede deel ook van toepassing zijn op de cohort studie onder intraveneuze drugsgebruikers.

WELKE PERSONEN WORDEN MEEGENOMEN IN DE ANALYSE?

Ten aanzien van de personen met se-nummers <2000, 6000-6999 en 8000-8999 die ooit seropositief getest zijn, kan er een voor de analyse relevant onderscheid gemaakt worden in de volgende groepen.

1. Personen die tijdens de follow-up geseroconverteerd zijn¹.
2. Personen die tussen oktober 1984 en april 1985 seropositief binnenkwamen. Hierbinnen is er weer een onderscheid te maken in
 - (a) Personen zonder verdere informatie uit eerdere periodes, die meteen weer de studie verlieten (code 23; de meesten wilden hun testuitslag niet weten). Eén van hen is bij 3(b) ondergebracht omdat hij in 1988 is teruggekeerd (nr. 376/8198).

¹Eén persoon, nr. 1095, is 2.8 jaar niet in follow-up geweest en in die periode geconverteerd.

- (b) Personen zonder verdere informatie uit eerdere periodes, die een tijd in follow-up zijn gebleven.
 - (c) Personen met informatie uit eerdere periodes, die een tijd in follow-up zijn gebleven².
 - (d) Personen, afkomstig uit de hepatitis B vaccinatie studie, die meteen weer de studie verlieten (code 23; de meesten wilden hun testuitslag niet weten)³.
3. Personen die na 1988 seropositief zijn binnengekomen. Binnen deze groep kan weer een onderscheid gemaakt worden in
- (a) Personen die een negatieve testuitslag hebben uit de periode voor intrede in het cohort⁴.
 - (b) Personen die reeds in 1985 in de studie hadden gezeten⁵.
 - (c) Personen die bij intrede niet vervolgd wilden worden (code 23), bijvoorbeeld omdat ze niet in aanmerking bleken te komen voor de AZT trial, en waarvan geen follow-up gegevens beschikbaar zijn.
 - (d) Overigen⁶.

Seroconverter studies beperken de analyse tot de personen uit groep 1.

Ten aanzien van survival studies waarin ook de data van prevalent positieven worden gebruikt, worden, naast de personen uit groep 1, ook de personen uit groep 2(b) en 2(d) meegenomen in de analyses. Omdat zij in het algemeen geen goede follow-up informatie hebben, worden personen uit groep 2(a) niet meegenomen, ook niet als bekend is dat ze later AIDS ontwikkelden (bijv. nummer 584). Personen uit groep 2(c) worden wel meegenomen, omdat zij wel goede follow-up gegevens hebben ten aanzien van AIDS en dood.

De keuze om personen uit de derde categorie mee te nemen in de analyses, is niet eenduidig te maken en zal afhangen van het soort onderzoek. De onzekerheid t.a.v. de datum van serocon-

²De nrs. 21 30 101 162 268 316 373 393 421 500 585 615 648 705 8301 8321 zijn afkomstig uit de hepatitis B vaccinatie studie, 372 heeft een seronegatieve test in het AMC in 1982, 481 639 1068 hebben een seropositieve test in het AMC een paar maanden voor intrede.

³Dit zijn de nrs. 17 18 45 120/6057 128/6040 325 375 503.

⁴De nrs. 6003 6065 6071 6073 6082 6088 6093 6102 6114 6115 6117 6125 6133 6155 8020 uit hepatitis B studie; de nrs. 6059 6090 6164 uit AMC.

⁵Twee personen, nrs. 51/6100 745/6104, zijn ruim zes jaar niet in follow-up geweest en in die periode geconverteerd. Verder nr. 376/8198, die reeds in 1985 seropositief was, maar toen zijn testuitslag niet wilde weten (code 23); de nrs. 120/6057 128/6040 behoren hier strikt genomen ook toe, maar deze hebben ook in de hepatitis B studie gezeten en staan daarom bij 2(d).

⁶Zeven personen hebben een eerdere seropositieve test, omdat zij seropositief testten in de periode dat alleen seronegatieven konden intreden. Dit zijn de nrs. 6006 6010 6012 6021 6022 6096 8314. De nrs. 6002 6008 6013 6020 6026 6051 6052 6060 6072 6095 6113 6136 6144 6146 8007 8014 8017 8018 8305 8314 8318 8323 8326 8331 8333 hebben een eerdere seropositieve test op het AMC. De nrs. 6059 6090 hebben een eerdere seronegatieve test op het AMC. De nr. 6164 heeft zowel een eerdere seronegatieve als een eerdere seropositieve test op het AMC. De nrs. 6136 6144 zijn eerder getest i.v.m. een primaire HIV-infectie en worden daarom niet als seroconverters meegenomen.

versie is erg groot⁷. Daarnaast moet er rekening gehouden worden met het feit dat dit mogelijk een selecte groep is in verband met de mogelijkheid van behandeling en dit tot een bias in de analyses kan leiden. Een aantal (zeer) snelle progressors zullen gemist zijn, omdat ze al AIDS hadden gekregen voor hun mogelijke intrede in de studie. Hiervoor kan, in principe althans, via left-truncation gecorrigeerd worden⁸. Aan de andere kant kunnen in deze groep personen die al enige HIV-gerelateerde klachten hadden oververtegenwoordigd zijn, bijvoorbeeld omdat ze afkwamen op de AZT-trials. De personen in groep 3(c) worden niet meegenomen in de analyses, omdat ze geen follow-up gegevens hebben.

In de database komt een variabele ENTRCAT (entrance category) die aangeeft tot welke van bovenstaande groepen iemand behoort. Verder komt er een variabele STUDY, die aangeeft aan welke studie(s) en protocollen iemand heeft deelgenomen.

De variabelen LNEG (date of last known seronegative test) en FPOS (date of first known seropositive test) zitten niet standaard in de database, maar kunnen eenvoudig worden berekend uit de testuitslagen van de bezoeken. Als er informatie is t.a.v. seroconversie voor intrede, dan wordt dit als een aparte event bijgehouden, inclusief waar deze informatie vandaan komt.

Naast deze groepen zijn er nog de 3000-nummers, de 7000-nummers en de 9000-nummers. Deze zijn hier niet in detail bekeken. De 9000-nummers (de restgroep) worden i.h.a. niet meegenomen omdat zij geen homoseksuele mannen zijn, de 7000-nummers worden niet meegenomen omdat zij via het AMC aangemeld zijn en een aantal van hen al verder in hun infectie verkeren. De 3000-nummers (jonge homoseksuelen) kunnen eventueel wel meegenomen worden in de analyses.

In een analyse met immunologische en virologische markers worden naast de personen uit groep 2(a) en groep 3(c), ook de personen uit groep 2(d) noodgedwongen weggelaten, omdat zij geen marker-informatie leveren, met uitzondering van de twee die later zijn teruggekeerd (120/6057 en 128/6040). De mogelijke problemen met de derde groep gelden hier waarschijnlijk in mindere mate. Het probleem van de hoge mate van onzekerheid t.a.v. de datum van seroconversie kan omzeild worden door de datum van eerste marker meting als nulpunt te nemen. Een voorwaarde is wel dat er een covariaat meegenomen wordt in de analyse die een goede representatie geeft voor de tijd sinds infectie. Bij CD4 aantal of percentage lijkt dit het geval te zijn; of dit voor andere markers ook geldt is nog niet bekend. Als bovendien de derde groep, na correctie voor CD4-aantal, een gelijk risico heeft op AIDS of dood als de andere twee groepen bij vergelijkbaar CD4-aantal (m.a.w. het is geen selecte groep meer), is het veilig om de derde groep mee te nemen in de analyses. Hetzelfde geldt voor de 7000-nummers.

⁷Een uitzondering vormen de nummers 6065 6117 6133 6155 8020 uit de hepatitis B studie; 6065 en 8020 zijn tijdens de vaccinatiestudie geconverteerd, de overige drie tijdens de follow-up na 1990, waarna ze tot de grote cohort-studie zijn toegetroten

⁸De praktische implementatie hiervan is echter niet eenvoudig wanneer dood als eindpunt genomen wordt, omdat AIDS i.p.v. dood het exclusiecriteria voor intrede in de studie was.

INFORMATIE MET BETREKKING TOT HET EINDPUNT

Het eindpunt van de studie – AIDS, dood door AIDS, dood (algemeen) – kan ofwel waargenomen worden, ofwel rechts gecensureerd zijn. Bij personen die nog steeds in follow-up zijn, is het duidelijk welke keuze er gemaakt moet worden. Personen die naar het AMC, het OLVG of het Slotervaart ziekenhuis zijn gegaan kunnen ook beschouwd worden als zijnde in follow-up, zodra de klinische follow-up gegevens gekoppeld zijn. Bij studies met markers worden personen rechts gecensureerd zodra zij de studie verlaten hebben.

Voor survival studies zonder markers is er in het algemeen extra informatie beschikbaar ten aanzien van de AIDS-status en dood na het verlaten van de studie⁹. Daarom worden voor de personen die de studie verlaten hebben de volgende variabelen toegevoegd:

Verhuisd naar het buitenland.

In leven en AIDS-vrij, plus woonplaats.

In leven en AIDS-status onbekend, plus woonplaats.

In leven met AIDS, plus woonplaats¹⁰.

Overleden, plus plaats van overlijden¹¹.

Indien van toepassing, hoort bij ieder van deze variabelen een datum van laatst bekende informatie met betrekking tot deze variabele. Daarnaast wordt de informatie toegevoegd hoe deze informatie verkregen is:

persoonlijk
via derden
AIDS-registratie
ziekenhuis
...

Informatie verkregen via de burgerlijke stand, indien aanwezig, wordt apart opgeslagen. Bewaard worden de laatste twee data van aanschrijven, en de informatie verstrekt door de burgerlijke stand op deze twee data.

Ten slotte wordt de meest recente datum van linken met de AIDS registratie bijgehouden.

Deze informatie is voldoende om voor iedere persoon tot een beslissing te komen. De volgende situaties kunnen optreden.

- **Op meest recente datum wonend in Nederland, en AIDS-status bekend.** Voor deze groep is het duidelijk wat gedaan moet worden.
- **Op meest recente datum wonend in Nederland, en AIDS-status onbekend.** Meestal zal deze informatie verkregen zijn via de burgerlijke stand, maar in principe kan dit ook via persoonlijk contact of via derden verkregen zijn. Personen die op deze datum in Amsterdam staan ingeschreven worden, vanwege vertraging in AIDS-rapportage, een half jaar

⁹Het afwezig zijn in de AIDS registratie wordt hier niet als informatie gezien

¹⁰Datum AIDS diagnose wordt al geregistreerd

¹¹Plus code die oorzaak overlijden aangeeft

voor de laatste datum van linken met de AIDS registratie verondersteld AIDS-vrij te zijn geweest (rechts gecensureerd). Als deze datum (linken minus 0.5 jaar) later is dan de meest recente datum van informatie dat de persoon in leven was, dan wordt de meest recente datum van informatie als datum AIDS-vrij genomen. Omdat elders in Nederland de AIDS registratie waarschijnlijk minder volledig is, worden personen die elders in Nederland wonen verondersteld AIDS-vrij te zijn geweest (rechts gecensureerd) twee jaar voor de meest recente datum van informatie. We veronderstellen hierbij dat de persoon binnen twee jaar na zijn AIDS diagnose zou zijn overleden; dit zal in de nabije toekomst waarschijnlijk bijgesteld moeten worden. Als er een latere datum dan twee jaar terug bekend is waarop deze persoon AIDS-vrij was (bijvoorbeeld via persoonlijk contact), dan wordt dit als datum van censurering genomen. Als de persoon eerder in Amsterdam woonde, en deze datum is minder dan twee jaar terug, dan wordt de eerste procedure toegepast, met als datum de datum van vertrek uit Amsterdam.

Wanneer dood (algemeen) of dood door AIDS het eindpunt van de analyse is worden deze personen rechts gecensureerd op de laatste datum van informatie dat de persoon in leven was.

- **Verhuisd naar het buitenland.** Bij deze personen wordt alle informatie die betrekking heeft op de periode na het vertrek naar het buitenland niet gebruikt, ook niet als dit een AIDS diagnose of overlijden betreft. Ten aanzien van de informatie vóór het vertrek naar het buitenland, gelden bovenstaande procedures, met als datum de datum van vertrek naar het buitenland. Merk op dat de reden van vertrek naar het buitenland gerelateerd kan zijn met ziekteprogressie, waardoor de censurering in werkelijkheid niet random is, en er dus een bias kan optreden¹².
- **Overlijden HIV gerelateerd, maar geen AIDS diagnose (code 885).** Deze krijgen een fictieve AIDS diagnose op de datum van overlijden (nrs. 100 300 526 650 705 1058 1161 6030 6076 6096 8301).
- **Overleden, oorzaak onbekend (code 884).** Het is onduidelijk wat hiermee gedaan moet worden. Gelukkig komt dit slechts sporadisch voor¹³. Deze personen worden in de analyses met AIDS of dood door AIDS als eindpunt rechts gecensureerd op de datum van hun laatste bezoek. Bij dood als eindpunt kunnen ze wel als event meegenomen worden.

¹²Op dit moment zijn de personen met de volgende se-nummers naar het buitenland vertrokken: de nrs. 203 420 491 1004(wereldreis) 6031 6070 6076 6122 (allen met code 21: reden uitval vertrek naar het buitenland), de nrs. 6041 6075 6103 8311 (allen met huidige code 555: niet gelukt op te sporen via de burgerlijke stand) en nummer 233 (geen van beide codes, zijn AIDS diagnose in het buitenland is tot nu toe ten onrechte meegenomen). Nummer 8311 had AIDS voor datum van vertrek naar het buitenland. Deze diagnose wordt meegenomen bij AIDS als eindpunt. Hij wordt rechts gecensureerd ten aanzien van overlijden op zijn datum van vertrek. Van nummer 547 is nog onduidelijk of deze een AIDS diagnose voor zijn vertrek heeft. Indien dit zo is, wordt hij behandeld als 8311, en anders als de overige nummers. De reden van vertrek naar het buitenland is bij de nrs. 6041 6075 547 en 8311 gerelateerd met ziekteprogressie.

¹³Dit zijn de nrs. 480 663 685, allen ongeveer zes jaar na hun laatste bezoek overleden. 480 in Leiden (heeft CD4 metingen vlak voor overlijden, tussen 20 en 160), 663 in Amsterdam, 685 in Den Haag.

- **Overleden, oorzaak niet HIV gerelateerd (code 887).** Deze worden rechts gecensureerd op datum van overlijden bij AIDS en dood door AIDS als eindpunt, terwijl zij een event hebben bij dood (algemeen) als eindpunt.

DE VARIABELEN IN DE ANALYSE.

Iedere survival analyse met seroconversie als nulpunt dient, naast allerlei cofactoren en markers, op zijn minst de volgende variabelen te bevatten:

- Datum van seroconversie (geschat).
- Datum waarop persoon “at risk” is. Dit is alleen van belang voor de personen uit groep 2 en 3, i.v.m. left truncation.
- Datum AIDS/dood/dood door AIDS of rechts censurering
- Indicator die aangeeft of iemand rechts gecensureerd is danwel een event heeft.

Alle seroconverters uit groep 1, behalve nummer 1095, hebben een periode tussen de laatste seronegatieve test en de eerste seropositieve test die minder is dan een jaar. Als datum van seroconversie wordt het midden tussen de laatste seronegatieve en de eerste seropositieve test genomen. Als nummer 1095 wordt meegenomen in de analyses, krijgt hij een geschatte datum van seroconversie op basis van zijn CD4-aantal bij de eerste positieve test.

Voor de personen uit groep 2 wordt de datum van seroconversie geschat op basis van hun CD4-aantal bij intrede. Een uitzondering vormen de personen uit de hepatitis B studie voor wie de periode tussen de laatste seronegatieve test en de eerste seropositieve test minder is dan een jaar of die geen CD4-metingen hebben omdat ze in 1985 meteen weer de studie verlieten. Als de breedte van het seroconversie interval minder is dan een jaar, wordt het midden tussen de laatste seronegatieve en de eerste seropositieve test genomen als datum van seroconversie. Bij personen zonder CD4-aantal wordt de datum van seroconversie geschat op basis van de seroconversieverdeling van het cohort. Om te corrigeren voor het feit dat een aantal snelle progressors gemist zijn worden personen uit groep twee at risk genomen vanaf de datum van intrede in de studie (bij AIDS als eindpunt, bij dood als eindpunt is een simpele maar niet geheel correcte procedure daar 1,5 jaar bij op te tellen).

Voor de personen uit groep 3 is op dit moment in de meeste gevallen geen redelijke schatting van de datum van seroconversie beschikbaar. De meest betrouwbare aanpak is het schatten op basis van de seroconversieverdeling van het cohort.

Bij de meeste personen uit groep 2, en zeker bij de personen uit groep 3, is de onzekerheid in de geschatte datum van seroconversie groot. Daarom geven standaard procedures te smalle betrouwbaarheidsintervallen.