

Reconstructiemogelijkheden van lange tijdreeksen van de doodsoorzakenstatistiek op basis van tijdreeksanalyse

Het CBS heeft sinds 1901 een statistiek van doodsoorzaken. Door herziening van de classificatie (International Classification of Diseases, inmiddels versie ICD-10) en door wijzigingen in de interpretatie van de classificatie ("codeerafspraken") zijn diverse gegevensbreuken in deze statistiek ontstaan. Van deze breuken is in het verleden geen consequente verslaglegging gemaakt.

De gegevensreeksen vanaf 1950 tot 2010 zijn onderzocht op breuken en op de mogelijkheid deze reeksen te reconstrueren zonder breuken. Hierbij worden gegevensbreuken in aanmerking genomen die ontstaan zijn bij de introductie van een nieuwe versie van de classificatie. In een uitgebreid vooronderzoek met 150 reeksen zijn ook significante breuken in tussenliggende jaren onderzocht. Er bleken onvoldoende plausibele verklaringen voorhanden voor deze breuken. Daarom zijn deze verder niet meegenomen in het onderzoek.

De gevolgde aanpak is een stapsgewijze combinatie van methoden uit diverse kennisgebieden. Het onderzoek is gebaseerd op de reeksen van de European Shortlist of causes of death* (ESL, august 1998). Deze lijst van 65 internationaal belangrijk geachte doodsoorzaken is uitgebreid met een viertal voor Nederland relevante specificaties, namelijk myocardinfarct, de ziekte van Parkinson, verdrinking en overige transportongevallen. Aanpassing van de codegroepen was hier en daar nodig vanwege de constructie van de tijdreeksen of vanwege lokale codeerpraktijken (bijv. 278.9 Aids). Publicatie naar de voor Nederland gangbare Beldo-lijst (Belangrijke doodsoorzaken) is mogelijk vanwege de kleine verschillen met de ESL, zoals uit de concordantietabel in de uitgebreide onderzoeksbeschrijving blijkt (zie onderstaande link). In de onderzochte periode zijn vijf versies van de ICD in gebruik geweest en er zijn diensgevolge 4 ICD-overgangen. De tijdreeksen (ICD-8, ICD-9, ICD-10) zijn naar het verleden uitgebreid voor de perioden waarin de ICD-6 of ICD-7 gebruikt werd. In deze tijdreeksen zijn met behulp van Ox/SsfPack-software significante breuken bepaald op overgangen van ICD-classificaties. Vervolgens is de grootte van elke breuk berekend en zijn door de software reparatievoorstellen gedaan.

De methode onderscheidt zich van andere tijdreeksmethoden doordat het criterium is gesteld dat er door de reconstructie geen overlidingsgevallen bij mogen komen of mogen verdwijnen. Met andere woorden: de som van de breuken in de reeksen moet 0 zijn; dit betekent dat bij een breuk alle sterfgevallen die uit een reeks verwijderd worden, aan andere reeksen worden toegevoegd.

Door deze voor de hand liggende voorwaarde kan een correctie die een bepaalde doodsoorzaakcategorie "netter" maakt in de tijd, het tijdsverloop van andere doodsoorzaken beïnvloeden, soms zelfs verstoren.

ICD-overgang	Aantal significante breuken	Mediane grootte van de breuken in % van aantal doden aan betreffende ziekte in breukjaar	Mediane grootte van de breuken in % van alle doden in betreffend breukjaar
6 – 7	16	11	0,16
7 – 8	27	16	0,22
8 – 9	25	16	0,15
9 – 10	9	21	0,20

Het aantal significante breuken in de 69 geselecteerde tijdreeksen op ICD-overgangen wordt in bovenstaande tabel gegeven. Behalve de significantie is ook de grootte van de breuken binnen de reeks en de impact van de breuken in de reeks ten aanzien van alle sterfgevallen in dat jaar van belang. Uit de laatste kolom blijkt dat het vooral kleinere reeksen betreft.

Het is echter niet uitvoerbaar gebleken om de tijdreeksen *samenhangend* te corrigeren. Bij reconstructie van een tijdreeks kunnen kleine modelwijzigingen namelijk resulteren in tientallen procenten verschil in resultaat. Om uit de reparatievoorstellen een verantwoorde keuze te maken, is het nodig te beschikken over additionele informatie met betrekking tot het absolute niveau van bepaalde tijdreeksen in het verleden. Hierover zou in principe meer - maar geen volledige - duidelijkheid kunnen worden geschapen door volledige hercodering van alle individuele doodsoorzaakrecords. Echter, indien in het verleden geen gemakkelijk onderscheid mogelijk was tussen twee doodsoorzaken die tegenwoordig wel onderscheiden worden, dan kan ook hercodering dit niet oplossen. Als voorbeeld voor zo'n onderscheid kan gelden: het onderscheid bij kwaadaardige tumoren tussen orgaan van oorsprong en uitzaaiingen elders.

Met de toegepaste tijdreeksanalyse kunnen significante breuken worden opgespoord en de grootte van breuken nauwkeurig worden bepaald. Het is daarmee een nuttige aanvulling op de inhoudelijke aanpak van breuken in tijdreeksen en kan het onderzoek naar trends in doodsoorzaken betrouwbaarder maken. Het repareren van de breuken in tijdreeksen van doodsoorzaken kan echter alleen wanneer op individuele oorzaken voldoende informatie aanwezig is. Het CBS levert voor onderzoekers een consistente berekening van bestaande breuken in een set van 69 tijdreeksen. Naast deze set zijn ook twee breukensets met doodsoorzaken van mannen en vrouwen afzonderlijk berekend. Met de nodige medische en demografische informatie en de breukgegevens voor deze ziekten is het daarmee mogelijk meer inzicht te krijgen in het verloop van de doodsoorzaak van keuze.

Een uitgebreide rapportage en de tabel met berekende breuken is te vinden in:

[Reconstruction possibilities of long-term time series of causes of death.](#)