



Centraal Bureau
voor de Statistiek

Webartikel 2014

Verschuivingen in de doodsoorzakenstatistiek bij de introductie van het automatisch coderen

Dr. P.P.M. Harteloh

12-12-2014 gepubliceerd op cbs.nl

1. Inleiding

Met ingang van het statistiekjaar 2013 codeert het Centraal Bureau voor de Statistiek (CBS) doodsoorzakenformulieren automatisch met behulp van het softwarepakket IRIS. Dit is een verschil met voorgaande jaren waarin doodsoorzakenformulieren handmatig werden verwerkt. De introductie van het automatisch coderen brengt verschuivingen in de doodsoorzakenstatistiek met zich mee. Om deze te documenteren en te analyseren is in een eerder stadium een zogenaamde 'bridgecoding study' verricht. Van het statistiekjaar 2009 zijn alle handmatig gecodeerde formulieren overgetypt en ingevoerd (in 2010–2011) en vervolgens (ook) door Iris (versie 4.2.0/2012) gecodeerd. Hierdoor wordt een vergelijking van handmatig en automatisch coderen per sterfgeval mogelijk. De uitkomsten van deze 'bridgecoding study' zijn door het CBS gebruikt bij de besluitvorming over de introductie van het automatisch coderen, waarop het statistiekjaar 2013 het eerste jaar is geworden waarbij IRIS in de productie van de doodsoorzakenstatistiek is ingezet. In deze publicatie geven we de resultaten van de 'bridgecoding study' weer voor de lijst met belangrijke doodsoorzaken (Beldo-lijst). Het stelt de lezer in staat de verwachte verschuivingen, gebaseerd op de 'bridgecoding study', te vergelijken met de waargenomen verschuiving bij de overgang van handmatig (2012) op automatisch coderen (2013). Een volledige rapportage van de 'bridgecoding study', waarin verschuivingen niet alleen in nog meer detail beschreven, maar ook geanalyseerd worden, verschijnt medio 2015.

2. 'Bridgecoding study'

Een 'bridgecoding study' is een onderzoek waarbij (in dit geval) het handmatig en automatisch coderen worden vergeleken door eenzelfde bestand met behulp van deze twee verschillende methoden te coderen. We hebben een dergelijk onderzoek uitgevoerd op het jaarbestand 2009 (n = 134 262). De belangrijkste bevindingen zijn:

- IRIS kan 65 procent van de doodsoorzakenformulieren volautomatisch coderen en van een onderliggende doodsoorzaak voorzien. Van het overblijvende gedeelte is 27 procent afgewezen, omdat het doodsoorzakenformulier niet volledig gecodeerd kon worden. Dit kan het gevolg zijn van spelfouten of het ontbreken van een term in het woordenboek. IRIS kan het formulier dan niet van een onderliggende doodsoorzaak kon voorzien. Ook kon 8 procent van de doodsoorzakenformulieren niet worden aangeboden aan IRIS (niet natuurlijke doodsoorzaken, maternale sterfte en doodgeborenen), omdat de programmatuur (nog) niet geschikt is deze te verwerken en van een onderliggende doodsoorzaak te voorzien.
- In 78 procent van de volautomatisch gecodeerde formulieren is er een exacte overeenstemming tussen IRIS en de handmatig gecodeerde formulieren wat betreft onderliggende doodsoorzaak (ICD-10, vier digits). Deze overeenstemming hangt samen met het detailniveau van de ICD-10 code. Op het drie digit niveau van CBS Statline publicaties (w.o. de Beldo-lijst) is de overeenstemming 85 procent en op ICD-10 hoofdstukniveau (één digit) 90 procent.

- De introductie van het automatisch coderen van doodsoorzaken brengt (eenmalige) verschuivingen in de statistiek met zich mee. Er is een significante toename van infectieziekten (met name van de niet-gespecificeerde diarree), van de endocriene aandoeningen (met name van uitdroging), van de psychische aandoeningen (met name van dementie), van de ziekten van zenuwstelsel of zintuigen (met name van de ziekte van Alzheimer) en als onderliggende doodsoorzaak. Er is een significante afname van ziekten van de luchtwegen (met name van longontstekingen), van ziekten van het spijsverteringsstelsel (met name van de niet-infectieuze diarree), van aandoeningen van nieren of urinewegen (met name van urineweginfecties) en van symptomen en slecht omschreven ziektebeelden (met name van ouderdom) als onderliggende doodsoorzaak.
- Het meest frequent treden discrepanties tussen handmatig en automatisch coderen op bij het coderen van: longontsteking, COPD, atriumfibrilleren, diabetes mellitus, hartfalen, metastasen, hartinfarct, urineweginfecties, de ziekte van Alzheimer en het Cerebrovasculair accident (CVA).

Verklaring voor deze discrepanties wordt gevonden in:

1. Verschil in het gehanteerde principe voor selectie van de onderliggende doodsoorzaak;
2. Verschil in codering van het vierde cijfer van een ICD-10 code;
3. Technische beperking van IRIS (d.w.z. in de beslistabellen of het 'lezen' van een doodsoorzakenformulier);
4. Het gebruik van additionele informatie bij coderen (vrije tekst/toelichtingen van arts: wel gelezen bij handmatig coderen, maar IRIS kan geen vrije tekst lezen);
5. Het hanteren van ICD-10 updates: niet bij handmatig, wel bij automatisch coderen.

3. Definities van de belangrijkste begrippen

IRIS: eigennaam van een computerprogramma voor het automatisch coderen van doodsoorzaken. Europese standaard. Ontwikkeld in eerste decennium van deze eeuw. Momenteel in beheer bij het Duitse DIMDI. Door middel van een woordenboek worden Nederlandse uitdrukkingen die voorkomen op een doodsoorzakenformulier vertaald in ICD-10 codes, en vinden er vervolgens de door de ICD-10 voorgeschreven modificaties of combinaties van codes en selectie van de onderliggende doodsoorzaak plaats. IRIS gebruikt daarbij de tabellen van het Amerikaanse Medical Mortality Data System (MMDS).

MMDS: Medical Mortality Data System. Softwarepakket voor het coderen van doodsoorzaken en het bepalen van de onderliggende doodsoorzaak. In ontwikkeling sinds de jaren zeventig van de vorige eeuw. In beheer bij het Amerikaanse National Centre of Health Statistics (NCHS). Het bestaat uit verschillende modules: tabellen voor het selecteren van de onderliggende doodsoorzaak (ACME), lijsten met controles (bijvoorbeeld op leeftijd of geslacht) van codes die in ACME worden ingevoerd (MICAR200), een omzetting van ICD codes in willekeurig gekozen getallen (entity reference numbers, ERN) om zo vertaling van medische uitdrukkingen onafhankelijk te maken van verschillende ICD versies (MICAR100), en een Engelstalige invoer module voor medische termen (SUPERMICAR). IRIS heeft de

Engelstalige invoer module van MMDS (SUPERMICAR) vervangen door software die het programma geschikt maakt voor gebruik in niet-Engelstalige landen.

Comparability ratio (CR): de frequentie van een ICD-10 code voor de onderliggende doodsoorzaak (x) in IRIS gedeeld door de frequentie van diezelfde code bij handmatig coderen.

$$CR_x = \frac{\frac{\text{Aantal } x \text{ in IRIS}}{n}}{\frac{\text{Aantal } x \text{ bij handmatig coderen}}{n}}$$

De CR geeft per ICD-10 code(range) een beeld van de te verwachten verschuivingen in statistiek als we overgaan *van* handmatig *op* automatisch coderen. We toetsen de CR's op significantie en beschrijven codes met een voorkomen van minimaal 100 per jaar bij automatisch of handmatig coderen. Een ongewijzigde verandering van voorkomen (CR = 1) betekent niet dat er geen verschuiving in codering van records is opgetreden. Een instroom van anders gecodeerde records kan immers worden gemaskeerd door een even grote uitstroom van anders gecodeerde records. Een CR geeft deze verschuivingen niet weer. Daartoe hanteren we het percentage perfecte overeenkomst (PCP) tussen automatisch en handmatig gecodeerde formulieren. Een CR geeft echter wel het gevolg voor de verwachte verandering in doodsoorzakenstatistiek weer, die de introductie van het automatisch coderen met zich meebrengt. Ook kan met behulp van de CR's het feitelijk voorkomen van een doodsoorzaak in 2013 (automatisch gecodeerd) worden teruggerekend naar het voorkomen van een doodsoorzaak in 2013 als ware deze handmatig gecodeerd, hetgeen onder meer van belang is voor het repareren van tijdreeksen.

Een perfect compatibility percentage (PCP): het percentage sterfgevallen dat bij automatisch en handmatig coderen precies dezelfde ICD-10 code (i) krijgt.

$$PCP_i = \frac{\text{Aantal (i in IRIS = i bij handmatig coderen)}}{\text{(Aantal i bij handmatig coderen)}} \times 100\%$$

Aangezien we overgaan *van* handmatig *naar* automatisch coderen is het aantal gevallen, gecodeerd bij handmatig coderen in de formule als noemer gekozen. De PCP kan echter ook gedefinieerd worden met het aantal sterfgevallen dat aan die code werd toegeschreven door IRIS als noemer. Omdat sommige codes handmatig beter worden gecodeerd dan automatisch of omgekeerd, is op voorhand niet aan te geven aan welke noemer de voorkeur moet worden gegeven. Daarom geven we de beide PCP's in tabel 4.1 weer. Bijvoorbeeld:

3.1 Kruistabel voor onderliggende doodsoorzaak I21.9 (hartinfarct)

Sterfgeval	IRIS (automatisch gecodeerd)		
	I21.9	Niet-I21.9	
Handmatig gecodeerd			
I21.9	a	b	a + b
Niet-I21.9	c	d	c + d
	a + c	b + d	n

$$CR_{121.9} = \frac{(a + c)}{(a + b)}$$

Twee mogelijkheden voor PCP:

PCP_{121.9} ten opzichte van handmatig coderen = $a/(a+b) \times 100\%$

PCP_{121.9} ten opzichte van automatisch coderen = $a/(a+c) \times 100\%$

4. Handmatige en automatische codering van de lijst met belangrijke doodsoorzaken (Beldo)

De tabel 4.1 toont een vergelijking van het handmatig en automatisch coderen voor de lijst met belangrijke doodsoorzaken (Beldo-lijst), zoals deze door het CBS op Statline wordt geplaatst. De vergelijking is gebaseerd op het aantal volautomatisch gecodeerde formulieren (ongeveer 65 procent van het jaarbestand, $n = 86\ 929$). Formulieren die IRIS om een of andere reden niet kan verwerken zijn niet in de vergelijking opgenomen. Dit is een belangrijk verschil met het jaarbestand 2013 waar de overblijvende formulieren met handmatige hulp van een codeur van een onderliggende doodsoorzaak wordt voorzien. In de tabel is per doodsoorzaak (genummerd volgens Beldo-lijst, naam met tussen haakjes de ICD-10 code(range)) aangegeven hoe vaak deze voorkomt bij respectievelijk automatisch (IRIS) en handmatig (HM) gecodeerde formulieren. De CR geeft (ceteris paribus) de verwachte verschuiving bij overgang van handmatig op automatisch coderen. Deze kan worden vergeleken met de waargenomen verschuiving, ofwel 2013 (automatisch) gedeeld door 2012 (handmatig). Deze waargenomen verschuiving wordt beïnvloed door meer factoren dan het verschil in methode enkel en alleen (zie: Harteloh et al., 2014, *'Het automatisch coderen van doodsoorzaken'*). Het 95%-betrouwbaarheidsinterval (BTI) kan bij deze vergelijking worden gebruikt.¹⁾ Bevat dit interval 1 dan is de CR van de 'bridgecoding study' niet significant en er met andere woorden geen significant verschil in voorkomen van de desbetreffende code bij handmatig of automatisch coderen. Als de ratio 2013/2012 buiten dit 95%-betrouwbaarheidsinterval is gelegen, wordt de verwachte CR, gevonden in de 'bridgecoding study', niet waargenomen, en zijn er met andere woorden andere factoren dan de overgang op automatisch coderen die het verschil tussen de jaarbestanden 2012 en 2013 bepalen. Zo heeft er in de 'bridgecoding study' geen handmatige controle en/of correctie van codes plaatsgevonden en is ook het portie formulieren dat door IRIS niet kon worden gecodeerd niet verder verwerkt.

¹⁾ SPSS: 95%-betrouwbaarheidsinterval voor riskratio's.

4.1 Verschil tussen handmatig en automatisch coderen voor de lijst met belangrijke doodsoorzaken (Beldo) met een voorkomen van meer dan 100 sterfgevallen per jaar

BELDO	IRIS (aantal)	HM (aantal)	CR ¹⁾	95%- BTI	2013/ 2012 ²⁾	PCP ten opzichte van HM	PCP ten opzichte van IRIS
1 Infectie ziekten (A00-B99)	1 805	1 226	1,47	1,37–1,58	1,43	82,0	55,5
1.1 Tuberculose (A15.0-A19.9, B90.0)	-	-	-	-	-	-	-
1.2 Nekkramp (A39)	-	-	-	-	-	-	-
1.3 Virale hepatitis (B15.0-B19.9)	-	-	-	-	-	-	-
1.4 AIDS (B20.0-B24.9)	-	-	-	-	-	-	-
1.5 overige infectieziekten	1 492	960	1,55	1,43–1,69	1,46	81,7	52,4
2 Nieuwvormingen (C00.0-D48.9)	28 845	29 071	0,99	0,98–1,00	0,99	97,5	97,8
2.1 Kwaadaardige Nieuwvormingen (KNV) (C00.0-C97.9)	28 096	28 584	0,98	0,97–1,00	0,98	96,7	97,9
2.1.1 KNV lip, mond en keel (C0.00-C14.9)	280	308	0,91	0,77–1,07	1,03	88,6	96,4
2.1.2 KNV slokdarm (C15.0-C15.9)	1 081	1 100	0,98	0,90–1,07	0,96	96,8	98,1
2.1.3 KNV maag (C16.0-C16.9)	1 047	1 059	0,99	0,91–1,08	0,98	97,0	97,7
2.1.4 KNV dikke darm (C18.0-C18.9)	2 512	2 608	0,96	0,91–1,02	0,92	94,7	97,9
2.1.5 KNV Endeldarm & anus (C19.0-C21.9)	717	748	0,96	0,87–1,06	0,93	96,5	96,4
2.1.6 KNV lever (C22.0-C22.8)	242	328	0,74	0,63–0,87	0,77	71,0	94,6
KNV lever (C22.0-C22.9)	448	465	0,96	0,85–1,10	-	91,8	94,4
2.1.7 KNV Galblaas & galwegen (C23.0-C24.9)	278	237	1,17	0,99–1,40	1,26	98,7	84,2
2.1.8 KNV pancreas (C25.0-C25.9)	1 754	1 791	0,98	0,92–1,05	0,96	96,9	98,8
2.1.9 KNV larynx (C32.0-C32.9)	118	117	1,01	0,78–1,30	1,00	93,2	91,5
2.1.10 KNV long (C33.0-C34.9)	7 472	7 632	0,98	0,95–1,01	1,00	96,6	98,2
2.1.11 Melanoom (C43.0-C43.9)	495	502	0,99	0,87–1,12	1,04	98,0	98,6
2.1.12 KNV borst (C50.0-C50.9)	2 366	2 397	0,99	0,93–1,04	0,99	95,6	96,5
2.1.13 KNV cervix (C53)	148	146	1,01	0,81–1,27	1,03	97,3	95,9
2.1.14 KNV baarmoeder (C54.0-C55.9)	241	246	0,98	0,82–1,17	1,02	93,9	95,4
2.1.15 KNV eierstok (C56)	709	729	0,97	0,88–1,08	0,98	96,7	99,2
2.1.16 KNV prostaat (C61)	1 895	1 912	0,99	0,93–1,05	0,99	94,8	95,6
2.1.17 KNV nier/urinewegen (C64.0-C64.9)	584	593	0,98	0,89–1,10	0,93	96,1	96,6
2.1.18 KNV blaas (C67.0-C67.9)	848	874	0,97	0,88–1,07	0,99	94,0	96,5
2.1.19 KNV bloed/lymfe (C81.0-C96.9)	2 010	1 929	1,04	0,98–1,11	1,03	96,9	92,7
2.1.20 Overige KNV	3 266	3 328	0,98	0,93–1,03	0,96	91,0	92,3
2.2 Overige nieuwvorming (D00.0-D48.9)	749	487	1,54	1,37–1,72	1,41	95,3	61,3
3 Ziekten van bloed (D50.0-D89.9)	296	258	1,15	0,97–1,36	1,05	67,4	58,4
4 Endocriene ziekten (E00.0-E99.9)	2 874	2 479	1,16	1,10–1,22	1,05	85,2	73,2
4.1 Diabetes (E10.0-E14.9)	2 072	2 030	1,02	0,96–1,08	1,03	85,6	83,5
4.2 Overige endocriene ziekten	795	442	1,80	1,60–2,02	1,11	77,4	43,0
5 Psychische stoornissen (F00.0-F99.9)	7 250	5 516	1,31	1,27–1,36	1,22	95,2	72,3
5.1 Alcohol (F10.0-F10.9)	265	132	2,01	1,63–2,47	2,13	90,9	44,9
5.2 Drugs	-	-	-	-	-	-	-
5.3 Overige psychische stoornissen	6 974	5 371	1,30	1,26–1,34	1,20	95,3	73,3
6 Ziekten van zenuwstelsel/zintuigen (G00.0-H95.9)	3 225	2 746	1,17	1,12–1,24	1,20	92,0	77,8
6.1 Meningitis (G00.0-G03.9)	-	-	-	-	-	-	-
6.2 Parkinson (G20-G21)	750	830	0,90	0,82–1,00	1,03	82,8	90,9
6.3 Overige ziekten zenuwstelsel/zintuigen	2 431	1 863	1,30	1,23–1,39	1,26	91,3	69,6
7 Ziekten van hart-vaatstelsel (I00-I99)	28 271	27 876	1,01	1,00–1,03	1,00	93,4	91,5
7.1 Ischemische hartziekten (I20.0-I25.9)	7 582	7 713	0,98	0,95–1,01	0,95	90,8	91,3
7.1.1 Acut hartinfarct (I21.0-I22.9)	5 190	5 351	0,97	0,94–1,01	0,92	89,6	91,2
7.1.2 Overige ischemische hartziekten (I20, I23-I25)	2 392	2 362	1,01	0,96–1,07	1,01	86,2	84,5
7.2 Overige hartziekten (I30-I33, I39-I52)	9 596	9 470	1,01	0,99–1,04	1,00	87,4	85,5
7.3 CVA (I60.0-I69.9)	7 173	6 486	1,11	1,07–1,14	1,11	94,4	84,9
7.4 Overige ziekten HV stelsel	3 866	4 207	0,92	0,88–0,96	0,95	77,9	88,9
8 Ziekten van ademhalingsorganen (AO) (J00.0-J99.9)	7 803	10 065	0,78	0,75–0,80	0,85	71,6	92,0
8.1 Griep/influenza (J10.0-J11.9)	-	-	-	-	-	-	-
8.2 Longontsteking (J12.0-J18.9)	2 131	3 931	0,54	0,51–0,57	0,64	49,3	90,6
8.3 COPD (J40.0-J47.9)	4 536	4 801	0,94	0,91–0,98	0,96	87,8	92,5
8.3.1 Astma (J45.0-J46.9)	-	-	-	-	-	-	-
8.3.2 Overige chronische luchtwegaandoeningen	4 466	4 762	0,94	0,90–0,98	0,95	87,2	0,3
8.4 Overige ziekten AO	1 026	1 254	0,82	0,75–0,89	0,94	65,9	92,7

4.1 Verschil tussen handmatig en automatisch coderen voor de lijst met belangrijke doodsoorzaken (Beldo) met een voorkomen van meer dan 100 sterfgevallen per jaar (slot)

BELDO	IRIS (aantal)	HM (aantal)	CR ¹⁾	95%- BTI	2013/ 2012 ²⁾	PCP ten opzichte van HM	PCP ten opzichte van IRIS
9 Ziekten van spijsverteringsstelsel (K00.0-K93.9)	2 238	2 625	0,85	0,81–0,90	0,85	76,5	88,9
9.1 Maagzweer (K25.0-K28.9)	–	–	–	–	–	–	–
9.2 Chr. Leveraandoening (K70, K73-K74)	343	362	0,95	0,82–1,10	0,93	83,7	86,9
9.2.1 Alcoholische leveraandoeningen (K70)	101	121	0,83	0,64–1,09	0,83	74,4	88,1
9.2.2 Overige Chr. Leveraandoeningen	242	241	1,00	0,84–1,20	1,03	87,6	85,5
9.3 Overige ziekten spijsverteringsstelsel	1 801	2 151	0,84	0,79–0,89	0,84	74,4	88,4
10 Ziekten van huid & bindweefsel (L00-L99)	114	162	0,70	0,55–0,89	0,77	49,4	70,2
11 Ziekten van spieren/bindweefsel (M00-M99)	327	295	1,11	0,95–1,30	1,17	76,3	68,8
11.1 Reumatoïde artritis (M05–06, M15–19)	–	–	–	–	–	–	–
11.2 Overige ziekten spieren/bindweefsel	195	207	0,94	0,77–1,15	1,19	67,1	71,3
12 Ziekten van nier/urineewegen & geslachtsorganen (N00-N99)	1 694	2 171	0,78	0,73–0,83	0,82	65,0	83,2
12.1 Ziekten van nier/urineewegen (N00-N20, N22-N29)	1 093	1 128	0,97	0,89–1,05	0,95	76,6	78,8
12.2 Overige ziekten nier/urineewegen & geslachtsorganen (N21, N30–99)	601	1 043	0,58	0,52–0,64	0,66	50,3	87,4
13 Complicaties van zwangerschap en bevalling (O00-O99)	–	–	–	–	–	–	–
14 Aandoeningen van perinatale periode (P00-P99)	–	–	–	–	–	–	–
15 Aangeboren afwijkingen (AA) (Q00-Q99)	167	166	1,01	0,81–1,25	1,01	84,9	83,2
15.1 AA van zenuwstelsel (Q00-Q07)	–	–	–	–	–	–	–
15.2 AA van hart en bloedvaten (Q20–28)	–	–	–	–	–	–	–
15.3 Overige AA (Q10–18, Q30–39)	144	140	1,03	0,82–1,30	0,99	87,9	84,0
16 Symptomen en onvolledig omschreven ziektebeelden (R00-R99)	2 006	2 242	0,89	0,84–0,95	1,03	77,7	86,3
16.1 Wiegendood (SIDS) R95	–	–	–	–	–	–	–
16.2 Onvolledig omschreven ziektebeelden en onbekende oorzaken (R96-R99)	393	399	0,98	0,86–1,13	1,15	79,7	79,6
16.3 Overige symptomen & onvolledig omschreven ziektebeelden (R00-R94)	1 611	1 841	0,88	0,82–0,94	0,91	76,7	87,3
Totaal	86 929	86 929	1,00	–	1,03	78,4	–

¹⁾ CR vet: significant verschil tussen automatisch en handmatig coderen in de 'bridgecoding study'.

²⁾ Ratio 2013/2012 cursief: de waargenomen verschuiving wordt mede bepaald door andere factoren dan de overgang op het automatisch coderen enkel en alleen.

Verklaring van tekens

.	Gegevens ontbreken
*	Voorlopig cijfer
**	Nader voorlopig cijfer
x	Geheim
–	Nihil
–	(Indien voorkomend tussen twee getallen) tot en met
0 (0,0)	Het getal is kleiner dan de helft van de gekozen eenheid
Niets (blank)	Een cijfer kan op logische gronden niet voorkomen
2013–2014	2013 tot en met 2014
2013/2014	Het gemiddelde over de jaren 2013 tot en met 2014
2013/'14	Oogstjaar, boekjaar, schooljaar enz., beginnend in 2013 en eindigend in 2014
2011/'12–2013/'14	Oogstjaar, boekjaar, enz., 2011/'12 tot en met 2013/'14

In geval van afronding kan het voorkomen dat het weergegeven totaal niet overeenstemt met de som van de getallen.

Colofon

Uitgever

Centraal Bureau voor de Statistiek
Henri Faasdreef 312, 2492 JP Den Haag
www.cbs.nl

Inlichtingen

Tel. 088 570 70 70, fax 070 337 59 94
Via contactformulier: www.cbs.nl/infoservice

© Centraal Bureau voor de Statistiek, Den Haag/Heerlen, 2014.
Verveelvoudigen is toegestaan, mits het CBS als bron wordt vermeld.