KCE REPORT 208As

Federaal Kenniscentrum voor de Gezondheidszorg
Centre Fédéral d'Expertise des Soins de Santé
Belgian Health Care Knowledge Centre

SYNTHESE

KOPPELING VAN DE PERMANENTE STEEKPROEF MET DE ZIEKENHUISGEGEVENS

HAALBAARHEIDS- EN REPRESENTATIVITEITSSTUDIE





Het Federaal Kenniscentrum voor de Gezondheidszorg

Het Federaal Kenniscentrum voor de Gezondheidszorg is een parastatale, opgericht door de programmawet (1) van 24 december 2002 (artikelen 259 tot 281) die onder de bevoegdheid valt van de Minister van Volksgezondheid en Sociale Zaken. Het Centrum is belast met het realiseren van beleidsondersteunende studies binnen de sector van de gezondheidszorg en de ziekteverzekering.

Raad van Bestuur

	Effectieve Leden	Plaatsvervangende Leden
Voorzitter	Pierre Gillet	
Leidend ambtenaar RIZIV (vice-voorzitter)	Jo De Cock	Benoît Collin
Voorzitter FOD Volksgezondheid (vice-voorzitter)	Dirk Cuypers	Christiaan Decoster
Voorzitter FOD Sociale Zekerheid (vice-voorzitter)	Frank Van Massenhove	Jan Bertels
Administrateur-generaal FAGG	Xavier De Cuyper	Greet Musch
Vertegenwoordigers Minister van Volksgezondheid	Bernard Lange	Brieuc Van Damme
	Bernard Vercruysse	Annick Poncé
Vertegenwoordigers Minister van Sociale Zaken	Lambert Stamatakis	Vinciane Quoidbach
	Ri De Ridder	Koen Vandewoude
Vertegenwoordigers Ministerraad	Jean-Noël Godin	Philippe Henry de Generet
	Daniël Devos	Wilfried Den Tandt
InterMutualistisch Agentschap	Michiel Callens	Frank De Smet
	Patrick Verertbruggen	Yolande Husden
	Xavier Brenez	Geert Messiaen
Beroepsverenigingen van de artsen	Marc Moens	Roland Lemye
	Jean-Pierre Baeyens	Rita Cuypers
Beroepsverenigingen van de verpleegkundigen	Michel Foulon	Ludo Meyers
	Myriam Hubinon	Olivier Thonon
Ziekenhuisfederaties	Johan Pauwels	Katrien Kesteloot
	Jean-Claude Praet	Pierre Smiets
Sociale partners	Rita Thys	Leo Neels
	Paul Palsterman	Celien Van Moerkerke
Kamer van Volksvertegenwoordigers	Lieve Wierinck	

Controle

Regeringscommissaris — Yves Roger

Directie

Algemeen Directeur — Raf Mertens
Adjunct Algemeen Directeur — Christian Léonard
Programmadirectie — Kristel De Gauquier

Contact

Federaal Kenniscentrum voor de Gezondheidszorg (KCE)
Doorbuilding (10e verdieping)
Kruidtuinlaan 55
B-1000 Brussel
België

T +32 [0]2 287 33 88
F +32 [0]2 287 33 85
info@kce.fgov.be
http://www.kce.fgov.be

KCE REPORT 208As
HEALTH SERVICES RESEARCH

Federaal Kenniscentrum voor de Gezondheidszorg
Centre Fédéral d'Expertise des Soins de Santé
Belgian Health Care Knowledge Centre

SYNTHESE

KOPPELING VAN DE PERMANENTE STEEKPROEF MET DE ZIEKENHUISGEGEVENS

HAALBAARHEIDS- EN REPRESENTATIVITEITSSTUDIE

CÉCILE CAMBERLIN, CÉCILE DUBOIS, TONIO DI ZINNO, RAGNA PRÉAL, JOERI GUILLAUME, REMI VAN DE REYD, CARINE VAN DE VOORDE, STEFAAN VAN DE SANDE, STEPHAN DEVRIESE

COLOFON

Titel:	Koppeling van de permanente steekproef met de ziekenhuisgegevens Haalbaarheids- en representativiteitsstudie – Synthese
Auteurs:	Cécile Camberlin (KCE), Cécile Dubois (KCE), Tonio Di Zinno (AIM – IMA), Ragna Préal (IMA – AIM), Joeri Guillaume (IMA – AIM), Remi Van de Reyd (IMA – AIM), Carine Van de Voorde (KCE), Stefaan Van De Sande (KCE), Stephan Devriese (KCE)
Reviewers:	Nathalie Swartenbroekx (KCE), France Vrijens (KCE)
Stakeholders:	Nathalie Bossuyt (Wetenschappelijk Instituut Volksgezondheid – Institut Scientifique de Santé Publique), Joëlle Carton (Cellule Technique de traitement de données relatives aux hôpitaux - Technische Cel voor de verwerking van de gegevens met betrekking tot de ziekenhuizen), Lut De Prins (RIZIV – INAMI), Ri De Ridder (RIZIV – INAMI), Pascal Meeus (Cellule Technique de traitement de données relatives aux hôpitaux - Technische Cel voor de verwerking van de gegevens met betrekking tot de ziekenhuizen), Ingrid Mertens (Technische Cel voor de verwerking van de gegevens met betrekking tot de ziekenhuizen – Cellule Technique de traitement de données relatives aux hôpitaux), Yves Parmentier (Cellule Technique de traitement de données relatives aux hôpitaux – Technische Cel voor de verwerking van de gegevens met betrekking tot de ziekenhuizen), Alain Quewet (SPF Santé publique – FOD Volksgezondheid), Guy Van Camp (FOD Sociale Zekerheid – SPF Sécurité sociale), Johan Van Der Heyden (Wetenschappelijk Instituut Volksgezondheid – Institut Scientifique de Santé Publique), Peter Willemé (Federaal Planbureau – Bureau fédéral du Plan)
Externe validatoren:	Dirk De Wachter (Vlaams Agentschap Zorg en Gezondheid), Niel Hens (Universiteit Hasselt), Véronique Tellier (Service public de Wallonie - DG05 - DSE - OWS)
Andere gemelde belangen:	Geen gemeld
Layout:	Ine Verhulst, Sophie Vaes

Disclaimer:

- **De externe experten werden geraadpleegd over een (preliminaire) versie van het wetenschappelijke rapport. Hun opmerkingen werden tijdens vergaderingen besproken. Zij zijn geen coauteur van het wetenschappelijke rapport en gingen niet noodzakelijk akkoord met de inhoud ervan.**
- **Vervolgens werd een (finale) versie aan de validatoren voorgelegd. De validatie van het rapport volgt uit een consensus of een meerderheidsstem tussen de validatoren. Zij zijn geen coauteur van het wetenschappelijke rapport en gingen niet noodzakelijk alle drie akkoord met de inhoud ervan.**
- **Tot slot werd dit rapport unaniem goedgekeurd door de Raad van Bestuur.**
- **Alleen het KCE is verantwoordelijk voor de eventuele resterende vergissingen of onvolledigheden alsook voor de aanbevelingen aan de overheid.**

Publicatiedatum:	17 oktober 2013
Domein:	Health Services Research (HSR)
MeSH:	Hospital records ; Medical record Linkage ; Sampling Studies ; Medical Informatics
NLM classificatie:	W 26.55.I4
Taal:	Nederlands
Formaat:	Adobe® PDF™ (A4)
Wettelijk depot:	D/2013/10.273/63

Copyright: De KCE-rapporten worden gepubliceerd onder de Licentie Creative Commons « by/nc/nd » http://kce.fgov.be/nl/content/de-copyrights-van-de-kce-rapporten.

Hoe refereren naar dit document? Camberlin C, Dubois C, Di Zinno T, Préal R, Guillaume J, Van de Reyd R, Van de Voorde C, Van de Sande S, Devriese S. Koppeling van de permanente steekproef met de ziekenhuisgegevens. Haalbaarheids- en representativiteitsstudie – Synthese. Health Services Research (HSR). Brussel: Federaal Kenniscentrum voor de Gezondheidszorg (KCE). 2013. KCE Reports 208As. D/2013/10.273/63.

Dit document is beschikbaar op de website van het Federaal Kenniscentrum voor de Gezondheidszorg.

VOORWOORD

Het systeem van betaling per prestatie – één van de pijlers van onze ziekteverzekering – deed al veel inkt vloeien, niet zelden kritische commentaren. Het heeft in elk geval het voordeel dat elke prestatie geregistreerd wordt. Vandaag willen we even die andere kant van de medaille belichten, en we doen dit vanuit het perspectief van ons eigen werk als adviesorgaan voor de beleidsmaker en ontwikkelaar van klinische richtlijnen. En het mag gezegd worden: weinig landen beschikken over zulk een gedetailleerd instrument om het zorggebruik van zijn burgers in kaart te brengen. De permanente steekproef van sociaal verzekerden met hun lange-termijn zorggebruik ontsluit op een vrij unieke wijze deze informatie. Een logische volgende stap is het toevoegen van diagnostische informatie. Met dit project wilden we verkennen of dit haalbaar en effectief bruikbaar is.

Het is een leerrijke ervaring geworden, en meteen een soort unicum in de KCE geschiedenis – nog nooit werden onze aanbevelingen zo direct bekrachtigd door het moeizame verloop van het project zelf. Dit vergt wat meer uitleg. In de plaats van de voorziene 12 maanden zal de studie uiteindelijk meer dan twee en een halve keer langer geduurd hebben, en de reden ligt voornamelijk in het omslachtige van de hele koppelingsprocedure, niet in het minst in de stappen die nodig waren om een machtiging te krijgen van het bevoegde sectoraal comité en deze ook getrouw uit te voeren. Uiteraard zijn al deze beschermingsmaatregelen volkomen gewettigd – maar ze maken wel dat projecten die dit soort gekoppelde gegevens nodig hebben erg duur en traag worden. Wat rechtstreeks onze aanbeveling onderbouwt en de noodzaak illustreert om – onder voorwaarden – een dergelijke koppeling permanent te maken.

Als we uiteindelijk tot een resultaat zijn gekomen, dan is het zeker mede dank zij de constructieve samenwerking met de collega's van het InterMutualistisch Agentschap, de Technische Cel en het eHealth Platform, en wij wensen hen dan ook te danken voor deze fijne synergie.

Christian LÉONARD
Adjunct Algemeen Directeur

Raf MERTENS
Algemeen Directeur

SAMENVATTING

Sinds 2005 heeft België een opmerkelijk instrument voor de analyse van de gezondheidszorggegevens: de Permanente Steekproef (EPS). Het is een extractie van een deel van de gegevens die Belgische verzekeringsinstellingen leveren aan het InterMutualistisch Agentschap (IMA). Deze steekproef bevat terugbetalingen van gezondheidsgerelateerde prestaties, een aantal persoonlijke bijdragen en socio-economische gegevens van een veertigste van de verzekerde populatie jonger dan 65 jaar en een twintigste van de verzekerde populatie van 65 jaar of ouder. De steekproeftrekking is gestratificeerd naar geslacht en leeftijd (volle geboortedatum). De Permanente Steekproef is een handig instrument, beschikbaar voor een aantal instellingen waaronder het InterMutualistisch Agentschap, het Rijksinstituut voor ziekte- en invaliditeitsverzekering (RIZIV), het Federaal Planbureau en het Federaal Kenniscentrum voor de Gezondheidszorg voor hun wetenschappelijke analyses van het gezondheidsbeleid. In tegenstelling tot de volledige gegevens van het InterMutualistisch Agentschap is de toegang niet onderworpen aan een langdurig proces dat de goedkeuring vereist van het Sectoraal Comité van de Sociale Zekerheid en Gezondheid van de Commissie voor de bescherming van de persoonlijke levenssfeer. Klinische gegevens (diagnoses en procedures) met betrekking tot verblijven in het ziekenhuis ontbreken in de Permanente Steekproef, maar zijn beschikbaar in de Minimale Ziekenhuisgegevens (MZG), beheerd door de Federale overheidsdienst Volksgezondheid, Veiligheid van de Voedselketen en Leefmilieu (FOD-VG). Deze klinische gegevens zijn nuttig om de resultaten van analyses aan te passen aan individuele risico's van patiënten, zoals comorbiditeit, of om patiënten te selecteren op basis van klinische kenmerken. De koppeling van deze twee complementaire databanken is zeer nuttig voor wetenschappelijk onderzoek, vooral als er een permanente koppeling zou zijn (jaarlijks herhaald).

In dit document worden de voorwaarden en de technische stappen nodig voor de koppeling, beschreven. De tweede doelstelling is de representativiteit van de beschikbare klinische variabelen te controleren na het koppelen van de Permanente Steekproef aan de medische gegevens gerelateerd aan ziekenhuisverblijven (klassieke ziekenhuisopnames en daghospitalisatie). Met andere woorden, in hoeverre kunnen de resultaten voor deze variabelen in de gekoppelde steekproef geëxtrapoleerd worden naar alle Belgische ziekenhuisverblijven?

Vóór de eigenlijke koppeling was het noodzakelijk om een procedure te ontwikkelen voor de uitwisseling van correspondentietabellen en gegevens tussen de verschillende actoren. Een reconstructiealgoritme voor ziekenhuisverblijven in de Permanente Steekproef werd gevalideerd. Van de verblijven geïdentificeerd in de Permanente Steekproef voor 2008-2009 werden 94% gekoppeld, wat als bevredigend werd beschouwd.

Statistische analyses op basis van de technieken ontleend aan de theorie van de steekproeftrekking werden uitgevoerd om de representativiteit van de volgende variabelen te testen: leeftijd, geslacht, voorkeursstatus, gewest, provincie, arrondissement, type ziekenhuisopname, graad van ernst (in de APR-DRG - All Patient Refined Diagnosis Related Group = groep van pathologiëen), hoofddiagnose en totaal bedrag van de terugbetalingen. Globaal komt de verdeling van elke variabele in de Permanente Steekproef overeen met de verdeling voor de nationale referentiepopulatie met een maximale foutenmarge van 1%. Op het lagere niveau (APR-DRG) zijn de verschillen in de verdeling van elke variabele kleiner dan 5% in 23% van de APR-DRG in de gekoppelde steekproef. 3% van de APR-DRG ontbreekt in de gekoppelde steekproef. Deze APR-DRG vertegenwoordigen niettemin 76% van de verblijven. Twee APR-DRG werden gedetailleerd uitgewerkt als voorbeeld. (1) In het geval van vaginale bevallingen zijn de resultaten representatief voor alle variabelen. (2) Voor de groep interventies op de knie en de onderste ledematen (met uitzondering van de voet), zijn de resultaten representatief behalve voor de gewest en de provincie, waarvoor verschillen tot 3% gevonden werden tussen de steekproef en de nationale gegevens. Op het globale niveau (alle verblijven samengenomen) is de representativiteit gewaarborgd, maar dit geldt niet altijd voor alle variabelen op lagere niveaus (APR-DRG en APR-DRG-graad van ernst). Daarom moet rekening gehouden worden met de meer gedetailleerde resultaten in functie van de specifieke onderzoeksvragen van elke studie.

SYNTHESE

1. ACHTERGROND EN DOELSTELLING VAN DE STUDIE

Sinds 2005 heeft ons land een opmerkelijk instrument voor de studie van het gezondheidsbeleid: de Permanente Steekproef. Deze database wordt samengesteld door het InterMutualistisch Agentschap, een associatie waarin alle verzekeringsinstellingen (VI) vertegenwoordigd zijn, en die sinds 2002 de gegevens van de VI samenbrengt met als doel deze gezondheidsgegevens te bestuderen op nationaal niveau. De EPS bevat de demografische, socio-economische gegevens en verzekeringsstatus alsook de eigen bijdrage en de terugbetalinggegevens (bv. medische dienstverlening, ziekenhuisforfaits, geneesmiddelen en medische hulpmiddelen zoals implantaten) van een fractie van de Belgische bevolking, representatief naar geslacht en leeftijd (1 op 40 voor degenen onder de 65 jaar en 1 op 20 voor de oudere groep).[1] Het doel van deze steekproef is het beschikbaar stellen van een handig hulpmiddel aan bepaalde instellingen[a], in tegenstelling tot alle gegevens van het IMA die alleen toegankelijk zijn na een uitgebreidere procedure. De Belgische regelgeving rond de bescherming van persoonsgegevens stelt dat elke overdracht van gegevens van het IMA volgens een strikte procedure moet verlopen waarbij een principiële machtiging is vereist van het Sectoraal Comité van de Sociale Zekerheid en Gezondheid.[4] Dit Comité gaat meer bepaald na of de indirecte identificatierisico's voor de patiënten en de zorgverleners tot een aanvaardbaar minimum zijn teruggebracht. Deze procedure is niet vereist voor de EPS.

[a] Deze instellingen zijn: RIZIV, FOD Volksgezondheid waaronder het Wetenschappelijk Instituut Volksgezondheid, de FOD Sociale Zekerheid, het KCE en het Federaal Planbureau.[2] Derden, zoals universitaire onderzoekers, regionale overheden of het Kankerregister kunnen een verzoek indienen tot toegang van een selectie van EPS-gegevens voor het beantwoorden van specifieke onderzoeksvragen. [3]

Wat betreft de ziekenhuisverblijven bevat de EPS geen diagnostische informatie. Deze informatie is dan wel weer terug te vinden in de Minimale Ziekenhuisgegevens (MZG) die sinds de jaren negentig door de Federale Overheidsdienst Volksgezondheid, Veiligheid van de Voedselketen en Leefmilieu (FOD-VG) worden beheerd. Naast bepaalde informatie, die nu al binnen de EPS (leeftijd, geslacht, woonplaats) beschikbaar is, bevatten de MZG klinische informatie over elk verblijf in een algemeen ziekenhuis in België (diagnoses, procedures en combinaties hiervan, onder de vorm van pathologiegroepen[b]). Aan de hand van die medische ziekenhuisgegevens kunnen groepen patiënten die aan bepaalde aandoeningen lijden, worden geanalyseerd of kunnen de analyseresultaten aan de individuele risico's van de patiënten worden aangepast. Om die redenen werden voor bepaalde wetenschappelijke studies van het KCE of het IMA, de IMA gegevens en de medische ziekenhuisgegevens eerder al aan elkaar gekoppeld.

Bij iedere nieuwe studie moet het administratieve proces voor het verlenen van machtigingen en het ter beschikking stellen van gegevens integraal worden doorlopen. De permanente koppeling van de EPS met de medische ziekenhuisgegevens zou dan ook bijzonder nuttig kunnen zijn. Het zou de toegang tot pathologiegegevens gekoppeld aan de IMA gegevens rond verstrekte zorgen in de ziekenhuisomgeving versnellen zonder dat een koppeling moet worden doorgevoerd van de volledige MZG en IMA databases of van ad hoc selecties ervan. De systematische machtiging van het Sectoraal Comité van de Sociale Zekerheid en Gezondheid, evenals de overdracht van gegevens uit de MZG en IMA databases zou dus niet langer bij elke studie zijn vereist.

Met dit gegeven voor ogen dient onderhavige studie van het IMA en het KCE een dubbel doel. In de eerste plaats is het de bedoeling om het MZG-EPS koppelingsproces en de technische implicaties (op IT en structureel vlak) nauwkeurig in kaart te brengen, ten behoeve van de gebruikers van EPS, in het bijzonder de instellingen die nu al rechtstreeks toegang hebben tot de EPS. In de tweede plaats, terwijl de EPS representatief is voor de leeftijd en het geslacht van de populatie van de Belgische verzekerden, ontbreekt alle informatie over de representativiteit van de klinische variabelen die de koppeling oplevert. Onderhavig rapport is er dan ook op gericht deze representativiteit na te gaan; het wil met andere woorden controleren of de resultaten die bij de analyse van de gekoppelde steekproef worden gevonden, kunnen worden veralgemeend tot het geheel van de Belgische ziekenhuisverblijven. Bij gezondheids-onderzoeken vormen het totale bedrag aan terugbetalingen in het kader van een ziekenhuisverblijf een vaak gebruikte variabele en derhalve werd de representativiteit van deze variabele eveneens nagegaan.

[b] APR-DRG (All Patient Refined Diagnosis Related Group)

2. METHODE

Een wettelijk kader voor de uitwisseling van correspondentietabellen tussen de gegevens van de verzekeringsinstellingen en deze van de MZG bestaat reeds geruime tijd. Sinds 1997 koppelt de Technische Cel (TCT) de door de FOD-VG overgemaakte MZG aan de terugbetalingsgegevens van de ziekenhuisverblijven, door de verzekeringsinstellingen aan het RIZIV overgemaakt (Anonieme Ziekenhuisverblijven (AZV) voor de klassieke hospitalisaties en Anonieme Daghospitalisaties (ADH) voor de dagverblijven). De TCT doet dit aan de hand van twee correspondentietabellen die zij ontvangt, enerzijds van de ziekenhuizen en anderzijds van de verzekeringsinstellingen. In het kader van dit project werd dan ook beslist om de reeds gekoppelde TCT gegevens te koppelen aan de EPS gegevens. Een schema voor de gegevensoverdrachten incluis bijhorende correspondentietabellen werd uitgewerkt waar de verzekeringsinstellingen, het Nationaal Intermutualistisch College (NIC), het eHealth platform, de Kruispuntbank van de Sociale Zekerheid, de FOD-VG, de Technische Cel en tot slot het IMA en het KCE bij werden betrokken. Dit schema beschrijft de uitwisseling van de correspondentietabellen, met dubbele hercoderingen van de patiëntidentificatoren, evenals de gegevensuitwisseling als dusdanig. De Technische Cel analyseerde het risico dat personen binnen kleine groepsentiteiten ("small cells") zouden kunnen worden geïdentificeerd (identificatie via frequentieanalyses op relevante variabelen). Het risico werd gereduceerd door het aantal gevraagde variabelen te beperken en door het hercoderen van bepaalde variabelen (bv. samenvoeging van leeftijden boven de honderd jaar). Het betrokken aantal verblijven bedroeg minder dan 0,01%. Die analyse moet vóór iedere studie, die betrekking heeft op EPS-TCT gekoppelde gegevens, worden overgedaan.

De TCT gegevens zijn per ziekenhuisverblijf gestructureerd, terwijl de EPS gegevens per prestatie zijn gestructureerd, zonder samenbundeling per verblijf. Aan de hand van een reconstructiealgoritme van de ziekenhuisverblijven in de EPS (op basis van de ziekenhuisforfaits) kon via het bepalen van begin- en einddatum van het verblijf dan ook in eerste instantie elk verblijf worden afgebakend. De koppeling op verblijfsniveau werd vervolgens op volgende in beide databases (TCT en EPS) aanwezige variabelen uitgevoerd: het pseudoniem van de patiënt, het erkenningsnummer van het ziekenhuis van het verblijf en de begin- en einddatum van het verblijf. De op die manier EPS-TCT gekoppelde verblijven omvatten zowel de klassieke ziekenhuisverblijven, als de daghospitalisaties voor de jaren 2008 en 2009.

Voor de "kosten" werden de terugbetalingen door verzekeringsinstellingen aan ziekenhuizen getotaliseerd. De berekening van het totaalbedrag van de terugbetalingen werd uitgevoerd door de Technische Cel vóór levering van de TCT-gegevens aan KCE en IMA en omvatte een herberekening van de dagprijs tegen 100% (inclusie van het deel van het Budget Financiële Middelen - BFM, niet per verblijf uitbetaald). Het doel bestond erin na te gaan of er tussen de steekproef en de nationale TCT-gegevens geen enkele vertekening bestond op het vlak van het totaal terugbetaalde bedrag.

De representativiteit werd getoetst op niveau van de verschillende APR-DRG (All Patient Refined Diagnosis Related Groups) en op het niveau van de APR-DRG-graad van ernst, omdat die niveaus van groepering van de verblijven doorgaans bij de studie van de medische ziekenhuisgegevens worden gebruikt. De globale resultaten worden integraal in het rapport voorgesteld samen met een beknopt overzicht op APR-DRG niveau. Om haalbaarheidsredenen werden slechts twee APR-DRG in detail toegelicht. Voor de afzonderlijke APR-DRG en voor elke APR-DRG-graad van ernst zijn de gedetailleerde resultaten evenwel beschikbaar.

Statistische analyse, gebaseerd op technieken van de steekproeftheorie, laat toe voor elke bestudeerde parameter (gemiddelde, mediaan, proportie, ...) een schattingsgrootheid en de precisie ervan (fout als gevolg van steekproefvariatie) te berekenen. De schattingsgrootheid, die samen met zijn 95% betrouwbaarheidsinterval wordt berekend op basis van de EPS-TCT gegevens, wordt vergeleken met zijn equivalent berekend op de totaliteit van de TCT gegevens (referentiepopulatie). De representativiteit is aangetoond wanneer het betrouwbaarheidsinterval van de schattingsgrootheid de waarde van de referentiepopulatie bevat.[5] De extrapolatie van het aantal verblijven in de steekproef werd ook vergeleken met het aantal verblijven van de referentiepopulatie. Die extrapolaties kunnen eventueel worden gebruikt bij epidemiologische analyses (berekening van de prevalentie of van de incidentie).

Variabelen en niveaus die werden bestudeerd:

- Leeftijd
- Geslacht
- Voorkeursstatuut
- Type verblijf
- Gewest, provincie en arrondissement
- APR-DRG
- Graad van ernst (van de APR-DRG)
- Hoofddiagnose
- Totaal bedrag van de terugbetalingen

3. RESULTATEN

Na het reconstrueren van de ziekenhuisverblijven in de EPS kon 93,8% van de EPS-verblijven worden gekoppeld. Dit percentage is vergelijkbaar met het koppelingspercentage van de EPS in voorgaande studies en kan als een succes worden beschouwd voor een première voor de koppeling over de gehele EPS. Het aantal gekoppelde verblijven bedroeg uiteindelijk 198 752 over twee jaar.

Globale analyse

Wanneer we alle ziekenhuisverblijven en alle APR-DRG samen nemen, dan stemt de verdeling van de individuele variabelen in de EPS-TCT overeen met deze van de TCT referentiepopulatie met een maximale foutenmarge van 1%. De gegevens van de gekoppelde EPS-TCT steekproef kunnen dan ook worden gebruikt om de parameters van de ziekenhuisgegevens (TCT) te schatten. Op dit globale niveau staat de “at random” steekproeftrekking in combinatie met een voldoende groot aantal verblijven inderdaad garant voor de representativiteit van de gekozen variabelen (geslacht, leeftijd, gewest, sociaal statuut, opnametype, ...).

Op “meso”-niveau was de representativiteit per instelling geen onderdeel van de scope; bijgevolg werd die ook niet specifiek onderzocht. Wel toonden de resultaten van de EPS-TCT-koppeling aan dat het koppelingspercentage van de verblijven op ziekenhuisniveau een belangrijke rol speelt. Dit percentage schommelt immers sterk tussen de instellingen onderling.

Analyse per APR-DRG

Voor respectievelijk 2008 en 2009 vertoonde in 22% en 23% van alle APR-DRG geen enkele van de geteste categoriale variabelen een afwijking van meer dan 5% tussen de verdeling binnen de EPS-TCT steekproef en deze binnen de nationale TCT gegevens. Deze APR-DRG vertegenwoordigen niettemin 76% van de verblijven. Voor de overige APR-DRG werd een verschil van meer dan 5% waargenomen voor de verdeling van minstens één variabele. Alle resultaten zijn beschikbaar per APR-DRG en per APR-DRG-graad van ernst. Ze worden hier niet weergegeven omwille van hun grote omvang, maar ze kunnen worden overgemaakt aan de instellingen met toegang tot de EPS of aan andere onderzoekers met een toegangsmachtiging. Niet alle APR-DRG zijn echter in de steekproef vertegenwoordigd: voor elk van beide jaren ontbreekt 3% van de APR-DRG. Voor de APR-DRG-graad van ernst ontbreken voor 33% (2008) en 32% (2009) van de APR-DRG minstens één graad van ernst op niveau van de steekproef t.o.v. nationaal.

Voor het onderzoek op APR-DRG niveau hebben we ervoor geopteerd om twee geslecteerde APR-DRG verder uit te diepen. Voor APR-DRG 560 "Vaginale bevalling" kon de representativiteit voor de verschillende variabelen (alle ernstgraden samen) worden aangetoond, en dus kunnen de EPS-TCT-gegevens van 2008-2009 worden gebruikt voor studies rond medisch begeleide vaginale bevallingen in België. De representativiteit voor de APR-DRG 313 "Ingrepen onderste extremiteiten knie en onderbeen, behalve voet" was voor het jaar 2009 globaal genomen bevredigend behalve voor het geografische niveau. Als we ervan uitgaan dat een foutenmarge van 3% onaanvaardbaar is, kan geen enkel resultaat per gewest of provincie binnen eenzelfde APR-DRG in eenzelfde steekproefjaar worden veralgemeend naar het geheel van de verblijven voor dit gewest of deze provincie in diezelfde APR-DRG.

4. DISCUSSIE

Haalbaarheid van de koppeling

De uitwisseling van de correspondentietabellen met een unieke versleuteling van de patiëntidentificator die de anonimiteit van de individuele gegevens garandeert, bestond al enkele jaren tussen de verschillende actoren (FOD–VG–TCT, RIZIV–TCT, VI–IMA). Dat er beroep kon worden gedaan op de expertise van de TCT voor de koppeling van de MZG-gegevens afkomstig van de FOD VG en van de AZV/ADH-gegevens afkomstig van het RIZIV, was een belangrijke troef in dit project. Het eHealth-platform speelde op zijn beurt de rol van intermediaire vertrouwenspartij (Trusted Third Party–TTP) in de uitwisseling van de gegevens door als tussenpersoon de identificatiesleutels te hercoderen.

Het reconstructiealgoritme van de ziekenhuisverblijven in de EPS werd opgebouwd op basis van de bestaande algoritmen die het IMA eerder in verschillende studies gebruikte. Hierdoor konden we dus de vruchten plukken van meerdere jaren ervaring binnen het IMA. De daaropvolgende succesvolle koppeling met de verblijven binnen de TCT-gegevens bevestigde bovendien de geldigheid van het gekozen algoritme. Hierbij dienden er bepaalde methodologische keuzes te worden gemaakt zoals het elimineren van bepaalde prestaties die niet konden beschouwd worden als representatief voor volwaardige ziekenhuisverblijven, met name de miniforfaits en het gebruik van de gipskamer. Voor specifieke studies zouden echter andere keuzes kunnen worden gemaakt waarbij een aanpassing van het reconstructiealgoritme zich dan opdringt.

Samengevat kon het voor de studie gebruikte reconstructiealgoritme van de ziekenhuisverblijven worden gevalideerd. Bovendien werd de koppelingsmethodologie verfijnd en werden de hiervoor noodzakelijke variabelen geïdentificeerd. Deze koppeling is bevredigend gebleken. Een koppeling is dus mogelijk en kan in de toekomst eventueel nog verbeterd worden door een analyse van de niet-gekoppelde verblijven.

Representativiteit

Onze representativiteitsanalyses hebben aangetoond dat de grootteordes globaal genomen worden gerespecteerd. Bovendien identifceerden zij zones waar die representativiteit minder goed was (subgroepen zoals per APR-DRG en APR-DRG-graad van ernst). Voortaan kunnen dus studies waarbij een analyse van de EPS-TCT-gegevens van 2008-2009 is vereist, met kennis van zaken worden uitgevoerd.

De in onderhavig project getoetste variabelen worden hoofdzakelijk gebruikt bij de studie van de ziekenhuisgegevens vanuit statistisch, financieel of epidemiologisch oogpunt. Er zijn daarnaast nog andere variabelen waarvoor de representativiteit niet werd getoetst. De toegepaste methodologie kan overigens ook voor die andere variabelen worden gebruikt, in functie van specifieke onderzoeksvragen. In de volgende drie gevallen moet die representativiteit absoluut worden getoetst:

- het gebruik van andere variabelen (zoals de opname op de spoeddiensten, of het inroepen van bepaalde diensten);
- de analyse op bijkomende detailniveaus van de eerder bestudeerde variabelen (subgroepen, kruising van variabelen). Zo werden subhoofdstukken van (hoofd)diagnoses niet bestudeerd, maar bundeling ervan zouden bijvoorbeeld potentieel relevante klinische groepen kunnen opleveren;
- het gebruik van andere jaren dan 2008-2009.

Bij iedere studie dient men alert te blijven en de specifieke eigenschappen van iedere gegevensbank niet uit het oog te verliezen. Zo bevat de EPS-TCT-steekproef geen gegevens van niet bij een VI aangesloten personen, omdat zij geen deel uitmaken van de EPS-populatie die immers door de VI worden geleverd. De verblijven met opname vóór het jaar van de koppeling (in casu 2008) zijn aanwezig in de TCT-gegevens, maar konden niet worden gekoppeld met EPS-informatie, die enkel de gegevens van het koppelingsjaar omvat. Dit kan belangrijk zijn voor seizoensgebonden analyses (begin van het jaar). Die laatste beperking zou in het geval van een permanent gekoppelde steekproef overigens niet langer bestaan. Bij koppeling van de eerdere jaren (2002-2007) moet rekening worden gehouden met de wijzigingen ingevoerd door de integratie van de kleine risico's van het stelsel van zelfstandigen in de verplichte verzekering sinds 2008. Zo ontbreken de gegevens over daghospitalisaties voor zelfstandigen (tot dan beschouwd als kleine risico's) in de EPS-gegevens van vóór 1 januari 2008. Bovendien zou de heterogene geografische spreiding van die zelfstandigen een impact kunnen hebben op de representativiteit van de EPS-TCT.

Kleine cellen (“small cells”) in de gekoppelde EPS-TCT gegevens

In dit project werd een kleine cel arbitrair gedefinieerd als tellende drie of minder personen. Naast het voor de hand liggende probleem van representativiteit, vormen deze kleine cellen een probleem in termen van de wet op de bescherming van persoonsgegevens (privacy wetgeving). [4] Het combineren van bepaalde variabelen draagt bij tot het risico van indirecte identificatie, in het bijzonder geslacht, leeftijd, woonplaats, etc.. maar ook pathologievariabelen en behandelingen (geneesmiddelen of procedures). Het risico van indirecte identificatie, alhoewel geldend voor de nationale gegevens, wordt verzwakt door de steekproeftrekking. Bovendien reduceert het samennemen van de categorieën opgelegd na de “small cell” analyses van de Technische Cel verder het risico.

Permanent karakter van de koppeling

Een permanente EPS-TCT-steekproef op basis van een jaarlijks aangevulde koppeling zou een beter beheer van middelen en een verhoogde haalbaarheid van bepaalde onderzoeksprojecten mogelijk maken. Extractie uit deze gekoppelde steekproef laat immers toe de haalbaarheid van een studie te toetsen alvorens de noodzakelijke procedure op te starten wanneer de MZG- en IMA-gegevens op nationaal niveau moeten worden gekoppeld. Dergelijk permanent karakter maakt bovendien longitudinale steekproefstudies mogelijk, waarbij de patiënt, dankzij een ongewijzigde identificatiesleutel, door de tijd heen kan worden gevolgd.

Bovendien laat de gekoppelde steekproef toe, naast de ziekenhuisgegevens, ook de terugbetalingsgegevens van geneesmiddelen en ambulante diensten vóór en na de ziekenhuisopname te onderzoeken, omdat deze gegevens deel uitmaken van de EPS. Op deze manier kunnen volledige zorgepisoden of zorgtrajecten wetenschappelijk onderzocht worden.

Terwijl voor iedere "ad-hoc" koppeling een machtiging van het Sectoraal Comité van de Sociale Zekerheid en Gezondheid nodig is, is er een formele wettelijke basis nodig voor een permanente koppeling.

Bij ieder nieuw initiatief is de eerste stap de duurste. De implementatie van het proces was bijzonder lang en complex, maar ze werkt. In dat opzicht kan een systematisering van de uitwisseling van correspondentietabellen tussen de verschillende actoren in het kader van een permanente EPS-TCT-steekproef de kwaliteit van de koppeling alleen maar verfijnen, aangezien de kwaliteit van de correspondentietabellen ieder jaar verbetert. De methode voor het testen van de representativiteit, uitgewerkt in deze studie, moet worden herhaald bij toekomstige koppelingen.

AANBEVELINGEN[c]

Aan de Minister van Sociale Zaken en Volksgezondheid:

- **Om "ad hoc" koppelingen te vermijden en de duur en kostprijs van studies te verminderen, en om longitudinale studies mogelijk te maken, zou er een permanente koppeling tussen de meest recente jaarlijkse EPS en TCT data moeten komen, met inbegrip van het ter beschikking stellen van de resultaten van de representativiteitstesten aan de gebruikers van de EPS-TCT steekproef. Hiertoe dienen de wettelijke en operationele aspecten spoedig geregeld te worden.**

[c] Alleen het KCE is verantwoordelijk voor de aanbevelingen.

REFERENTIES

1. Wet van 27 december 2005 houdende diverse bepalingen, B.S. 30 december 2005.
2. RIZIV. Permanente steekproef (gezondheidszorg) [Web page]. 2010 [cited July 30, 2013]. Available from: http://www.inami.fgov.be/information/nl/sampling/
3. IMA. De Permanente Steekproef [Web page]. 2012 [cited 6 August 2013]. Available from: http://www.aim-ima.be/nl/imaweb/DT/content/imaweb/datas/eps/eps_introduction.html
4. Wet van 8 december 1992 tot bescherming van de persoonlijke levenssfeer ten opzichte van de verwerking van persoonsgegevens, B.S. 18 maart 1993. Available from: http://www.ejustice.just.fgov.be/wet/wet.htm
5. Shqiponje T, Hutton D, Davies P, Adnett N, Coe R. Competition Within Schools: Representativeness of Yellis Sample Schools in a Study of Subject Enrollment of 14-16 Year Olds. In: Institute for Education Policy Research; 2004. p. 17.