



Big data in de gezondheidszorg

Definitie, toepassingen
en uitdagingen

Colofon

Auteur

Stefan Ottenheijm

Redactie

Karin Oost

Advies

Johan Krijgsman
Lucien Engelen
Tom van de Belt
Roel Schenk

Vormgeving

Media&More

Nictiz / TrendITION

Bezoekadres

Oude Middenweg 55
2491 AC Den Haag
T 070 – 3173450
F 070 – 3207437
www.trendition.nl

Postadres

Postbus 19121
2500 CC Den Haag

Copyright

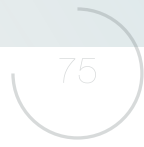
Nictiz, december 2015

Inhoudsopgave

1. Inleiding	4
2. Big Data: een definitie	8
2.1 Van data naar wijsheid	9
2.2 Meer, sneller, beter	9
2.3 Complexer = waardevoller?	10
2.4 Een definitie	11
3. Big data: toepassing in de gezondheidszorg	12
3.1 Kansen voor onderzoek	14
3.2 Voorkomen is beter dan genezen	15
3.3 Kwalitatieve, veilige en betaalbare zorg	17
3.4 Gepersonaliseerde zorg	18
3.5 Populatiemanagement en volksgezondheid	18
3.6 Fraude opsporing	19
4. Big data: veel kansen, veel uitdagingen	20
4.1 Technologie	21
4.2 Standaardisatie	21
4.1 Toegang tot data	21
4.2 Privacy	22
5. Tot slot	23
Geraadpleegde bronnen	23

1.

Inleiding



De wijze waarop Snow de gegevens combineerde om gebeurtenissen te verklaren en nieuwe, onverwachte inzichten op te doen, is een voorbeeld van big data 'avant la lettre'. Ruim 150 jaar later proberen we deze principes nog steeds toe te passen in de gezondheidszorg. Een sector die sinds jaar en dag veel gegevens verzamelt, maar er nog lang niet in slaagt om er alle kennis uit te halen die er potentieel in verscholen ligt. Een groeiend aanbod van data uit talloze (nieuwe) bronnen en vooruitgang in analyse-technologieën zorgen er samen voor dat we deze potentie steeds beter kunnen benutten.

In vrijwel elke sector speelt het begrip 'big data' al een rol en vragen beleidsmakers, bestuurders, directeuren en managers zich af wat ze er mee kunnen. En in veel sectoren wordt er al met succes mee gewerkt. Data worden op grote schaal verzameld en geanalyseerd om klantprofielen op te stellen, het productaanbod te optimaliseren, aanbiedingen op maat te kunnen doen en dienstverlening persoonlijker te maken. Zo kunnen winkels door koopgedrag van hun klanten en data van sociale media bijvoorbeeld beter bepalen waar de interesses van hun klanten liggen in een bepaalde tijd van het jaar. Hun productaanbod, voorraad en prijzen kunnen ze daarop afstemmen. Over het algemeen zorgen deze ontwikkelingen voor een hogere klanttevredenheid, een efficiënter gebruik van middelen en/of meer omzet. Maar gaan deze kansen ook op voor de gezondheidszorg?

De Nationale DenkTank, bestaande uit jonge academici, vindt van wel. In haar onderzoek naar big data in het publieke domein uit 2014 dichten zij grote kansen toe aan de inzet van big data in de gezondheidszorg. Sterker nog, volgens de auteurs is er geen sector die er zoveel bij te winnen heeft als de gezondheidszorg. Omdat de zorg efficiënter, meer op maat, met minder fouten en daardoor goedkoper kan worden geleverd. Als vruchtbare basis voor die potentie dient, naast de bergen data die zorgprofessionals verzamelen, de groeiende hoeveelheid data die burgers (al dan niet ongemerkt) verzamelen door middel van wearables, sensoren, e-mails, tweets, foto's, Facebook-berichten etcetera. Deze hoeveelheid neemt exponentieel toe.

Big data in de gezondheidszorg zit nu nog vooral in de hype fase: er wordt veel over gesproken, maar de praktijk is nog weerbarstig. Of zoals Dan Ariely van Duke University het verwoordde: "Big data is like teenage sex: everyone talks about it, nobody

really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it." Het Rathenau Instituut maakte een iets minder pikante vergelijking, en stelt dat big data zich nog in de 'goudkoortsfase' bevindt: "iedereen ziet wat in Big Data, maar tegelijkertijd is onduidelijk waar de waarde precies zit en hoe dit ontgonnen kan worden" (Rathenau Instituut, 2015).

Deze publicatie geeft een inkijkje in de wereld van big data in de gezondheidszorg: wat verstaan we eronder, wat is de potentie en wat doen we al?



“Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it.”



2. Big Data: een definitie



87

62

75

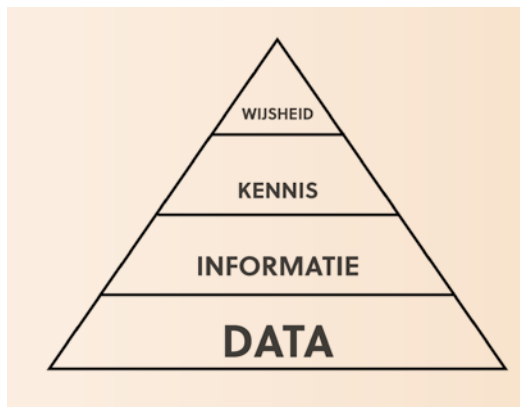
2.

Big Data: een definitie

2.1 Van data naar wijsheid

Om misverstanden te voorkomen is het belangrijk om allereerst het verschil aan te geven tussen de begrippen data en informatie, kennis en wijsheid. Alavi & Leidner (2001) hebben deze begrippen verhelderd in hun zogenaamde kennispiramide, zie figuur 2. Het begrip data vormt in dit model de onderste laag en bestaat uit onbewerkte feitelijke gegevens¹ die de werkelijkheid representeren. Informatie, een niveau hoger, is zinvol verwerkte data waaraan een interpretatie is gegeven. Kennis bestaat uit informatie die zinvol verwerkt is en wordt gezien als informatie op geaggregeerd niveau die tot een verandering in gedrag of verstand kan leiden (Huysman, 2006). Tenslotte wordt de top van de piramide gevormd door het begrip wijsheid, bestaande uit zinvol verwerkte kennis.

¹ De termen 'data' en 'gegevens' worden in deze verkenning regelmatig door elkaar heen gebruikt.

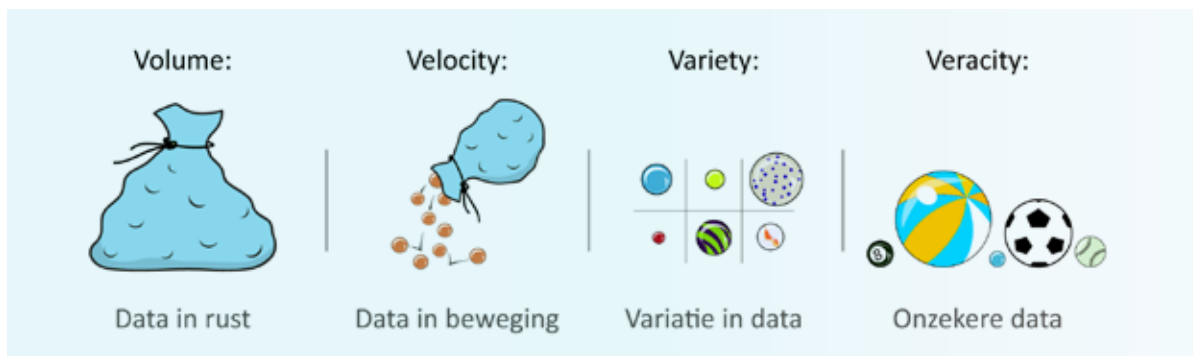


Figuur 2. Het onderscheid tussen data, informatie, kennis en wijsheid (Alavi & Leidner, 2001).

2.2 Meer, sneller, beter

Big data heeft betrekking op meer dan alleen grote hoeveelheden data. Een vaak gebruikte omschrijving van big data die een praktische benadering geeft, is afkomstig van Mark Beyer en Douglas Laney (2012) van onderzoeksbureau Gartner en bestaat uit de zogenaamde drie V's: *volume*, *velocity* en *variety*. Het eerste kenmerk is de grote hoeveelheden aan data (*volume*). Ter illustratie: iedere dag creëren we 2,5 triljoen gigabyte aan data. Dat neemt zo snel toe, dat 90% van alle wereldwijde data in de laatste

twee jaar is gecreëerd. Ook de snelheid waarmee data gegenereerd worden, veranderen en zich verspreiden is kenmerkend (*velocity*). Zo worden per seconde alleen al 2,5 miljoen mails en 10.000 twitter berichten verstuurd. Daarnaast zijn de verschillende vormen van data ook van belang (*variety*). Naast documenten vol cijfers en letters zijn er tal van andere vormen van data zoals audio, video en foto's. Aan deze drie V's wordt tegenwoordig in de literatuur ook nog een belangrijke vierde V toegevoegd: *veracity* (*waarheidsgetrouwheid*). Oftewel de



Figuur 3. Eigenschappen van big data (Schenk, 2015).

betrouwbaarheid of integriteit van al deze data waar analyses op los gelaten worden. Van veel data is immers onbekend door wie en op welke manier het gegenereerd is.

De verschillende kenmerken van verschillende soorten data komen ook tot uiting in de databronnen voor big data analyses in de gezondheidszorg. Volgens het Institute for Health Technology Transformation (2013) zijn er vijf verschillende categorieën of informatiestromen te onderscheiden:

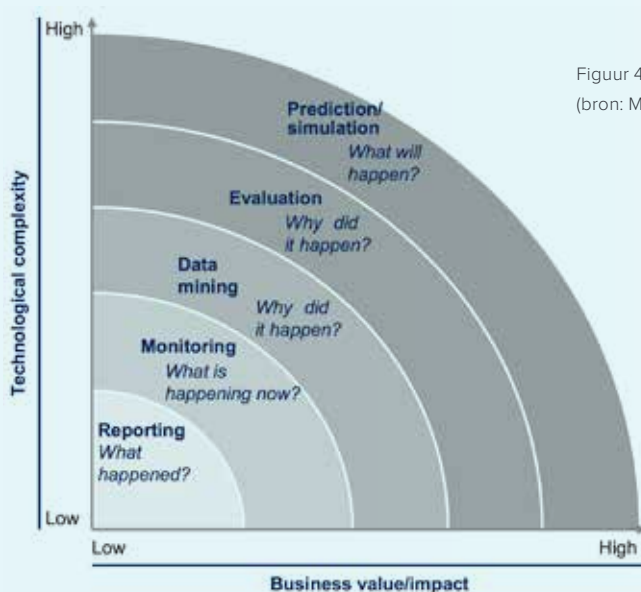
1. Internet en social media: klik- en surfgedrag op internet en interactie data van sociale media als Facebook, Twitter, LinkedIn etc.
2. Machine 2 machine: data uit sensoren en andere meetapparatuur, bijvoorbeeld de data uit een medicatiedispenser.
3. Transactie data: declaratiegegevens en andere gegevens uit de financiële informatiestromen in de gezondheidszorg.
4. Biometrische data: röntgenfoto's en ander beeldmateriaal, vingerafdrukken, genetische informatie, irisscans etc.
5. 'Human-generated' data: door mensen ingevoerde gegevens uit elektronische patiënten dossiers (EPD), aantekeningen, e-mails en papieren documenten.

De data in deze bronnen zijn daarnaast onder te verdelen in gestructureerde en ongestructureerde data. Data zijn gestructureerd als deze al op enige wijze in categorieën of volgens bepaalde logische structuren is verwerkt. Bijvoorbeeld declaraties met een vaste structuur die een informatiesysteem produceert of de gegevens die een bloeddrukmeter of weegschaal steeds op dezelfde wijze genereert en presenteert. Geschat wordt dat circa twintig procent van alle data gestructureerd zijn. Tachtig procent van alle data zijn dus ongestructureerd (Grimes, 2008). Ongestructureerde data zijn bijvoorbeeld medische beelden afkomstig uit de cardiologie of radiologie en (hand) geschreven verslagen in natuurlijke tekst. Voor wat betreft de omvang van medische beelden als ongestructureerde data: dertig (!) procent van de totale wereldwijde gegevensopslag wordt gebruikt voor medische beelden.

2.3 Complexer = waardevoller?

McKinsey heeft in de studie "The 'big data' revolution in healthcare" (2013) beschreven welke typen van big data-analyses er te onderscheiden zijn. Zij introduceren een model dat bestaat uit twee assen waarlangs deze typen worden afgezet: enerzijds de toegevoegde waarde en/of impact en anderzijds de technologische complexiteit. Volgens de auteurs neemt met de verandering van data, via informatie en kennis naar wijsheid, niet alleen de toegevoegde waarde toe, maar ook de technologische complexiteit om die waarde te bereiken. Concreet maakt dit model onderscheid tussen de volgende opeenvolgende niveaus:

- *Rapporteren*: wat is er gebeurd?
Voorbeelden op dit niveau bestaan onder andere uit 'eenvoudige' databanken die gebruikt kunnen worden voor het maken van (interne) rapportages of die antwoord kunnen geven op vragen als hoeveel operaties in een ziekenhuis hebben plaatsgevonden in een bepaalde periode. Dit betreft vaak data die door middel van zogenaamde dashboards als zinvolle informatie kunnen worden gepresenteerd (aan bijvoorbeeld bestuurders, zorgprofessionals of zorgconsumenten).
- *Monitoren*: wat gebeurt er nu?
Voorbeelden bij dit niveau kunnen op die uit het bovenstaande niveau (rapporteren) lijken, maar zijn uitgebreider omdat ze het ook mogelijk maken om de actuele situatie te vergelijken met een benchmark of een gewenste situatie. Er wordt gebruik gemaakt van zowel recente als real-time data. Bijvoorbeeld om te waarschuwen voor contra-indicaties bij het voorschrijven of verstrekken van medicatie. Andere voorbeelden zijn het monitoren van (complex) voorraadbeheer en de OK-planning en -bezetting in een ziekenhuis.
- *Data mining en evalueren*: waarom heeft een gebeurtenis plaatsgevonden?
Data mining en evalueren betreffen het
- *Voorspellen en simuleren*: wat gaat er gebeuren?
Het meest complexe niveau van dit model betreft het ultieme gebruik van big data: op basis van data kunnen voorspellen wat er in de toekomst mogelijk gaat gebeuren. Een (bekend) voorbeeld van technologie die op dit niveau data analyseert met als doel verwerking tot kennis en wijsheid is Watson van IBM. Deze supercomputer kan van specifieke patiënten persoonlijke data verwerken en vergelijken met onder andere de meest recente wetenschappelijke vakliteratuur, vergelijkbare situaties bij andere patiënten en het effect van hun behandeling om uiteindelijk met een individueel afgestemde diagnose, en eventueel behandeladvies, te komen en daarbij een betrouwbaarheidsmarge van die voorspelling te geven.



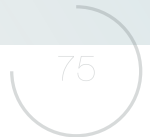
Figuur 4. Gradaties in technologische complexiteit (bron: McKinsey Business Technology Office, 2013)

2.4 Een definitie

Een alomvattende en algemeen geaccepteerde definitie van big data is moeilijk te geven. Het begrip wordt niet alleen gebruikt om de hoeveelheid, of complexiteit, van de data aan te duiden. In de praktijk wordt de term vooral gebruikt als containerbegrip dat de ontwikkeling van het verkrijgen van nieuwe kennis en wijsheid uit deze data betreft. Daarom stellen wij de volgende definitie voor: *big data refereert naar de mogelijkheid om gebeurtenissen te volgen, verklaren en voorspellen door het slim combineren en analyseren van complexe datasets uit verschillende bronnen.*

3.

Toepassing in de gezondheidszorg



3.

Big Data: Toepassing in de gezondheidszorg

Sectoren als retail, horeca, toerisme, energie en financiële instellingen zetten al big data-analyses in om hun dienstverlening te verbeteren. Organisaties in deze branches proberen allemaal de beschikbare data te gebruiken om het gedrag en de wensen van klanten in kaart te brengen, om risico's in te schatten en te voorspellen en om hun diensten en prijzen beter hierbij te laten aansluiten. Maar op welke manier kan dit in de gezondheidszorg en waar gebeurt dit al?



Binnen de gezondheidszorg worden big data-analyses voor verschillende doeleinden uitgevoerd. Daarbij kan grofweg een tweedeling gemaakt worden voor wat betreft de toepassing. Bij data die gebruikt worden voor de bedrijfsvoering wordt gesproken over *business intelligence of business analytics*. Dit zijn bijvoorbeeld digitale dashboards met inzichtelijk gepresenteerde informatie voor managers en bestuurders in het ziekenhuis. Data die ingezet worden ten behoeve van de zorg aan patiënten en wetenschappelijk onderzoek vallen onder de noemer *medical intelligence*, waarbij gedacht kan worden aan beslissingsondersteunende software voor de arts of misschien zelfs al het doen van voorspellende analyses over behandeluitkomsten.

Hieronder volgen enkele toepassingsgebieden aan de hand van voorbeelden uit binnen- en buitenland. Deze lopen uiteen van al in de praktijk gerealiseerde tot conceptuele toepassingen.

3.1 Kansen voor onderzoek

Het gebruik van big data kan de wetenschap vooruit helpen als het wordt ingezet bij het inrichten en uitvoeren van onderzoek. In klassiek wetenschappelijk onderzoek wordt eerst een onderzoeksvraag geformuleerd waarna er steekproefsgewijs data wordt verzameld om de hypothese te toetsen. Met behulp van de doorgaans relatief bescheiden hoeveelheden data, zoeken onderzoekers

bijvoorbeeld naar een causaal verband tussen twee eigenschappen. Bij modern 'big data-onderzoek' gaat dit andersom. Eerst wordt er data verzameld, vaak veel grotere gegevensverzamelingen, om daar vervolgens in te zoeken naar verwachte en onverwachte verbanden. Daardoor worden bevindingen gedaan die men met steekproeven nooit gevonden had.

Ook in de geneeskunde kan het zo werken, bijvoorbeeld bij onderzoek naar de werking van medicijnen. Bij een traditioneel opgezet onderzoek worden medicijnen getest en toegelaten op basis van de gemiddelde resultaten op een selecte groep patiënten. In die onderzoeksgroep zitten patiënten waarbij het medicijn wel effect heeft, maar doorgaans ook patiënten waarbij dit niet zo is. Toch wordt het medicijn uiteindelijk voorgeschreven aan alle patiënten, indien de onderzoeksgroep er gemiddeld baat bij had. Andersom zijn er ook medicijnen die de markt niet halen omdat een onderzoeksgroep er gemiddeld geen baat bij had, terwijl individuele patiënten dit misschien wel hadden. In deze traditionele zin van onderzoek doen, wordt er nog te weinig gedaan met deze verschillen tussen individuele patiënten. Het gebruik van meer data die de specifieke kenmerken van iedere individuele deelnemende patiënt in kaart brengen, kan helpen om beter inzicht te krijgen in waarom medicijnen bij bepaalde patiënten wel werken en bij anderen niet.

Data verzamelen binnen Life Lines (Universitair Medisch Centrum Groningen)

LifeLines is in december 2006 gestart als grootschalig bevolkingsonderzoek onder ruim 165.000 inwoners van Noord-Nederland (Groningen, Friesland en Drenthe). In dit onderzoeksprogramma worden deelnemers uit drie generaties minimaal dertig jaar gevolgd om inzicht te krijgen in hoe mensen gezond ouder kunnen worden en in de factoren die van belang zijn bij het ontstaan en verloop van chronische ziekten. De deelnemers worden eens in de ongeveer vijf jaar uitgenodigd voor een screening. Hierbij worden metingen en tests gedaan (zoals bloeddruk, gewicht, lengte, hartfunctie) en wordt er biomateriaal verzameld (zoals bloed, urine, DNA, hoofdhaar). De uitkomsten van de metingen en tests worden teruggekoppeld naar de deelnemer en zijn of haar huisarts. Daarnaast ontvangen de deelnemers iedere één tot twee jaar een uitgebreide vragenlijst met vragen over onder andere gezondheid, leefstijl en voedingsgewoonten. Centraal staat de vraag waarom de ene patiënt al relatief vroeg in het leven een chronische ziekte ontwikkelt en de andere tot op hoge leeftijd vitaal blijft. Het uitgangspunt van de onderzoekers is dat het ontstaan van chronische aandoeningen als astma, diabetes of nierziekten een complex samenspel van factoren is. De invloed van de verschillende factoren en de manier waarop ze op elkaar inwerken, zijn alleen te begrijpen door de gezondheid van een grote groep mensen uit meerdere generaties langdurig te volgen. De uitkomsten van LifeLines moeten leiden tot het sneller vaststellen van ziekte, het vinden van nieuwe behandelingen of zelfs tot het voorkomen van verschillende chronische aandoeningen (UMCG).

Meer informatie over Life Lines: <https://www.lifelines.nl>

Nieuwe initiatieven als 23andme, Apple's ResearchKit, Google Genomics en Patientslikeme laten zien hoe informatie- en internettechnologie potentie hebben om de mogelijkheden van het doen van onderzoek te verruimen. Hoewel de genoemde initiatieven van elkaar verschillen in opzet en doel, hebben ze gemeen dat ze gedreven worden door data die grote hoeveelheden individuele gebruikers aanleveren. Door gebruik te maken van de diensten van deze organisaties, kunnen wetenschappers in een mum van tijd veel participanten voor hun onderzoek werven en sneller en beter onderzoek doen. Veel grote farmaceutische bedrijven willen dan ook samenwerken met deze initiatieven.

3.2 Voorkomen is beter dan genezen

Preventie is ook een toepassingsgebied waarin de potentiële waarde van big data naar voren komt. Met de introductie en de verdere ontwikkeling van smartphones is er een wereld open gegaan voor het verzamelen van informatie over gedrag, leefstijl en gezondheid. Handige applicaties veranderen de mobiele telefoon in een stappenteller, slaapmonitor of medicatiebewaker. Een andere ontwikkeling is de opkomst van wearables van bedrijven als Fitbit, Jawbone en Withings waarmee gebruikers hun gedrag en bepaalde gezondheidswaarden kunnen meten, opslaan, delen en vergelijken met anderen (Nictiz, 2014).

Waar we deze informatie nu met name gebruiken om onze eigen voortgang te

monitoren en te vergelijken met anderen die dezelfde wearable of applicatie gebruiken, kan deze informatie ook worden gebruikt in de gezondheidszorg. Zo kan iemand verzamelde informatie over voeding en beweging direct delen met een behandeld arts. Als informatie bovendien gedeeld wordt voor (wetenschappelijk) onderzoek, kunnen onderzoekers iets zeggen over de staat gezondheid van populaties en kan het gebruikt worden om problemen en aandoeningen te herkennen voordat ze zich openbaren.

Een voorbeeld hiervan is het Amerikaanse bedrijf Propeller Health dat zich richt op ziekte management voor astma- en COPD-patiënten. Patiënten plaatsen hiervoor een sensor op hun inhalator, die registreert waar, wanneer en hoe vaak de inhalator gebruikt wordt. Door deze gegevens te combineren met zo'n veertig overige informatiebronnen zoals weersomstandigheden, verkeer, luchtkwaliteit of pollen in de lucht, kan het bedrijf patiënten en zorgaanbieders helpen met de uitkomsten van slimme analyses van deze data. Bijvoorbeeld door risicoprofielen te maken en risicogebieden in kaart te brengen, trends weer te geven en te voorspellen welke patiënten het meeste risico lopen. Patiënten kunnen een astma-aanval voor zijn doordat ze tijdig worden gewaarschuwd. Het bedrijf is onlangs gaan samenwerken met de stad Louisville in Kentucky (Verenigde Staten) en haar lokale apotheken om risicofactoren in de stad in kaart te brengen.

3.3 **Kwalitatieve, veilige en betaalbare zorg**

Naast de potentie om problemen en aandoeningen op te sporen voordat ze zich openbaren als ziekte, biedt big data ook kansen voor het leveren van betere zorg aan patiënten. Dat kan zich zowel uiten in betere uitkomsten van zorg als in de toegenomen betaalbaarheid ervan.

Een interessant voorbeeld dat de potentie hiervan goed schetst, is te vinden in de Verenigde Staten bij de zorgorganisatie Aurora Health Care. Deze organisatie bestaat uit vijftien ziekenhuizen, 185 klinieken en ruim tachtig apotheken. De ruim dertigduizend medewerkers binnen deze zorginstellingen creëren dagelijks een enorme hoeveelheid data. Een aantal jaren geleden ontstond het voornemen om deze data te gebruiken om beter inzicht te krijgen in de geleverde kwaliteit, om deze vervolgens te verbeteren en kosten te besparen.

Om dat te bereiken is 'Smart Chart' ontwikkeld (Forbes, 2012), een systeem dat alle data die in de laatste tien jaar binnen de organisatie verzameld waren, in één datawarehouse plaatst. Hierbij valt te denken aan facturatiegegevens, medische gegevens, laboratoriumwaarden, apothekersinformatie en de gegevens over afspraken. Met behulp van al deze databronnen kan inzichtelijk worden gemaakt hoe de kwaliteit van de geleverde zorg en de kosten ervan verschillen

tussen alle Aurora's instellingen. Daarbij kan Aurora zowel naar individuele patiënten als naar groepen patiënten met dezelfde aandoening kijken. Dit geeft antwoorden op vragen als: hoe lang blijven patiënten in de zorginstelling, hoe ziet het behandeltraject eruit, welke complicaties hebben ze, hoeveel heropnames zijn er? Zo ontstaat een beeld van de optimale uitkomsten voor patiënten en de behandeling die daarbij hoort. Door gebruik te maken van alle beschikbare data en de bijna real time data-analyse, kan Aurora de uitkomsten van verschillende behandelingen zelfs vrij nauwkeurig voorspellen. Een ander voordeel is dat artsen en onderzoekers geholpen worden om aan de juiste patiënten de juiste medicatie voor te schrijven.

Door deze resultaten te vergelijken met nationale benchmarks, krijgen de instellingen en hun afdelingen en artsen bovendien inzicht in hun eigen presteren. Dit draagt ook bij aan de constante zoektocht naar verbetering binnen de gehele organisatie. Het heeft erin geresulteerd dat Aurora Health Care een van de best scorende zorgaanbieders in de Verenigde Staten is. Sinds de introductie van de nieuwe werkwijze met data als hulpmiddel, is de afgelopen jaren het aantal heropnames met tien procent gedaald, wat resulteerde een miljoenenbesparing. Het benutten van de kennis die in de data verscholen ligt, helpt zo de zorg steeds weer te verbeteren en de kosten te drukken.

Big data in chirurgie

In het Radboudumc zijn de eerste stappen van big data zichtbaar in de diagnose-planning-behandel-evaluatie-cyclus bij chirurgie. Voor een volledige en adequate diagnose worden verscheidene beeldvormingstechnieken gecombineerd. Doel is om een virtuele patiënt te bouwen die een zo goed mogelijke afspiegeling geeft van de klinische en anatomische realiteit, met onder andere het gebruik van 3D-fotografie. Door kenmerken van patiënten te vergelijken met een steeds grotere groep eerdere patiënten, wordt een zo nauwkeurig mogelijke diagnose gesteld. Op basis van deze inzichten wordt een behandelplan opgesteld waarna de patiënt *virtueel* geopereerd wordt. Een simulatie van de postoperatieve situatie wordt gegenereerd op basis van de normale operatiestappen, die nu virtueel worden uitgevoerd. Een belangrijk voordeel is dat de arts op basis van big data analyses een verwachte uitkomst kan geven aan de patiënt, zodat deze volledig geïnformeerd en met realistische verwachtingen geholpen kan worden. Deze planningsstap zorgt voor een betere voorbereiding en duidelijker behandelplan preoperatief, waardoor de operatie sneller en voorspelbaarder verloopt.

Ook al is het optimale behandelplan in goed overleg opgesteld, de ideale uitkomst kan pas bereikt worden als dit plan ook precies kan worden gevolgd door de chirurg in de operatiekamer. Gebruik van technieken als navigatie geassisteerde chirurgie, 3D-geprinte boor- en zaagmallen en augmented reality zorgen voor een soepeler verloop van de operatie en een daling van het aantal complicaties en secundaire correcties. Kwantitatieve analyse en kwalitatieve evaluatie door de arts versterken elkaar vervolgens wederzijds in het evaluatieproces. Dit zorgt ervoor dat de volgende diagnose, planning en operatie met dezelfde middelen nog beter gaan dan bij de vorige patiënt, juist omdat de verkregen informatie aan de big databank wordt toegevoegd. De database van virtuele afspiegelingen van patiënten groeit dankzij deze cirkel op dagelijkse basis in het Radboudumc en zal in de toekomst vitale gegevens aanleveren voor steeds beter wordende big data ondersteuning.

3.4 Gepersonaliseerde zorg

Een andere kans van big data is de mogelijkheid om zorg te personaliseren, meer op maat voor het individu aan te bieden. 'Personalized healthcare' (gepersonaliseerde zorg) is een belofte die beetje bij beetje gestalte krijgt. Het gaat erom dat unieke kenmerken van individuen bepalend zijn bij het stellen van diagnoses, het voorspellen van het verloop van ziekten en het kiezen van behandelingen. Bijvoorbeeld de aanwezigheid van risicofactoren en specifieke genetische en metabolische kenmerken van de patiënt. Door gebruik te maken van informatie die bekend is over de uitkomsten van behandelingen van patiënten met vergelijkbare kenmerken, kunnen behandelingen gekozen worden die voor deze patiënt de beste uitkomsten bieden. Zo kan men ziektes op een veel effectievere manier behandelen dan voorheen.

Center for Personalized Cancer Treatment

Bij de behandeling van kanker werd tot voor kort voornamelijk uitgegaan van de 'gemiddelde kankerpatiënt'. In de praktijk betekent dit dat de behandeling van bijvoorbeeld longkanker of borstkanker bij een deel van de patiënten aanslaat en bij een deel van de patiënten niet. Alle patiënten krijgen echter te maken met bijwerkingen van de behandeling.

Elke (soort) kanker heeft echter bij elke patiënt weer andere kenmerken. Om kanker zo goed mogelijk te kunnen behandelen, lijkt het daarom logischer uit te gaan van elk individueel geval en niet van die 'gemiddelde kankerpatiënt' die eigenlijk niet bestaat. Geen algemene benadering, maar een gepersonaliseerde benadering dus. Het Center for Personalized Cancer Treatment (CPCT) probeert dit te bereiken door aan het begin van een behandeltraject het genetische materiaal van de kankercellen, het DNA, te analyseren en er een profiel van te maken. Door deze informatie te koppelen aan andere gegevens over uitkomsten van behandelingen en effecten van specifieke medicatie, kan de uitkomst van een specifieke behandeling voor een patiënt voorspeld worden. Patiënten krijgen in de toekomst zo een behandeling op maat en hoeven zo mogelijk niet meer blootgesteld te worden aan een inefficiënte behandeling.

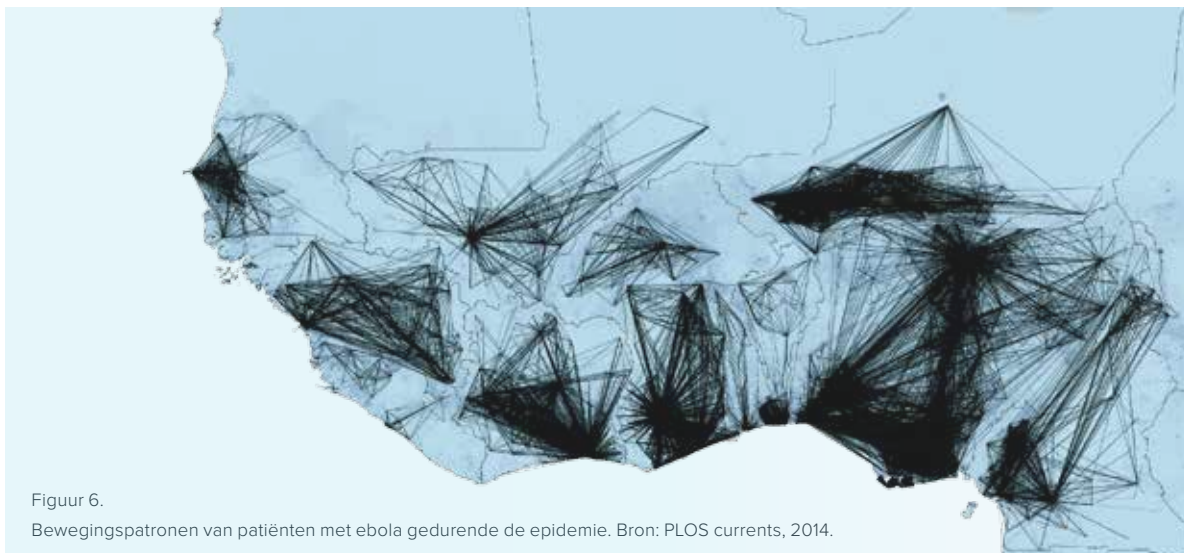
Meer informatie: <http://www.cpct.nl>

3.5 Populatiemanagement en volksgezondheid

Publieke gezondheidsinstellingen en organisaties uit het sociale domein kunnen ook baat hebben bij het combineren en analyseren van data uit de verschillende bronnen die zij beheren. Zo kunnen bijvoorbeeld verschillende risicogroepen in populaties onderscheiden worden. Door deze informatie te koppelen met gegevens over zorggebruik, kan de zorgprevalentie voor een bepaald gebied voorspeld worden. Een voorbeeld van de toepassing van dergelijke software is het zogenaamde Adjusted Case Mix (ACG) systeem dat is ontwikkeld door de John Hopkins universiteit uit Baltimore (VS). Het wordt gebruikt om een populatie onder te verdelen op geslacht, leeftijd, geografische locatie en andere kenmerken om de verschillen in gezondheidstoestand en zorggebruik te bepalen. Zo kan de software bijvoorbeeld berekenen hoeveel zorg een bepaalde bevolkingsgroep nodig heeft. De kennis die voorkomt uit dit soort voorspellende analyses kan gebruikt worden om het juiste zorgaanbod in de regio te organiseren en om preventieprogramma's op de juiste plek aan te bieden. Ook zorgverzekeraars gebruiken dergelijke voorspellingen ter onderbouwing van hun zorginkoop.

Organisaties die verantwoordelijk zijn voor de volksgezondheid kunnen ook baat hebben bij het gebruik van big data. Bijvoorbeeld bij het volgen van de verspreiding van infectieziekten en het analyseren van ziektepatronen om de publieke gezondheid te verbeteren en de reactiesnelheid bij incidenten te vergroten. Zo werd in 2014 in West-Afrika tijdens de ebola-uitbraak met behulp van data van telecomproviders in kaart gebracht hoe groepen mensen zich verplaatsten in de regio. Dit hielp om te bepalen waar hulpverlening het beste ingezet kon worden en wat de meest effectieve manier was om reisverboden in te stellen om verdere verspreiding te voorkomen.

Een stap verder zou het kunnen gaan als data uit andere bronnen (medische dossiers, sociale media, waarnemingen van hulpverleners ter plekke, gegevens over medicatieleveringen en verbruik) gecombineerd kunnen worden om te bepalen of reisverboden, educatieprogramma's en behandelingen ook daadwerkelijk werken. Of om zelfs te voorspellen welke plekken het meeste risico lopen voor snelle verspreiding van infectieziekten.



Figuur 6.
Bewegingspatronen van patiënten met ebola gedurende de epidemie. Bron: PLOS currents, 2014.

3.6 Fraude opsporing

Big data brengt ook hoge verwachtingen met zich mee om de opsporing van financiële onregelmatigheden en fraude te verbeteren. Uit onderzoek van de Nederlandse Zorgautoriteit (NZa, 2014) bleek dat medisch specialisten, huisartsen, verpleeginstellingen en andere zorgaanbieders in 2012 voor zo'n achthonderd miljoen euro aan opmerkelijke declaraties hebben ingediend. Ze factureerden bijvoorbeeld opvallend veel meer uren dan collega's, of lieten behandelingen stelselmatig doorlopen tot het moment waarop ze een hogere vergoeding konden krijgen. Het betreft hier dus een schatting van onregelmatigheden, waarvan een deel mogelijk door fraude veroorzaakt wordt. Schattingen over de financiële schade door fraude in de gezondheidszorg in Nederland lopen uiteen van tientallen miljoenen tot miljarden euro's per jaar (PWC, 2013).

In de enorme hoeveelheden data die organisaties uit de zorg jaarlijks aanleveren bij

zorgverzekeraars, kunnen met behulp van slimme analyses onregelmatigheden en fraude mogelijk sneller en beter herkend worden. In de Verenigde Staten hebben de Centers for Medicare and Medicaid Services (CMS), een federale verzekeringsorganisatie, dit tot op zekere hoogte al gedaan. Tot voor kort werden declaraties van patiënten digitaal en handmatig vergeleken met aangeleverde documentatie van de zorgprofessionals. Nu vergelijkt het 'Fraude Prevention System' declaratiepatronen met profielen van fraudeurs, worden zorgaanbieders eruit gepikt die meer uren declareren dan dat er in een dag passen en wordt er met behulp van gegevens uit sociale netwerken rekening gehouden met samenwerkingspartners van eerdere fraudeurs. Op deze manier worden honderden miljoenen dollars per jaar aan fraude opgespoord, significant meer dan de jaren voordat dit nieuwe systeem in werking werd gesteld. De kosten van het systeem worden momenteel met een ratio van één op tien terugverdiend.

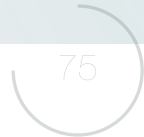
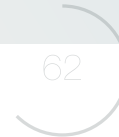
Gezondheidseffecten van luchtverontreiniging

Het Rijksinstituut voor Volksgezondheid en Milieu (RIVM) heeft samen met de Universiteit Utrecht onderzoek gedaan naar de effecten van luchtverontreiniging op de gezondheid. Met behulp van nieuwe onderzoeksmethoden is van een onderzoekscohort van zeven miljoen Nederlandse inwoners gedurende zeven jaar door het CBS geregistreerd wie er is overleden. Deze gegevens zijn anoniem gekoppeld aan andere data van het CBS, zoals leeftijd, geslacht, burgerlijke staat en woonadres. Op basis van de (versleutelde) woonadressen zijn vervolgens gegevens over fijn stof en stikstofdioxide aan de woonadressen van de zeven miljoen deelnemers gekoppeld. De resultaten laten zien dat zowel de hoeveelheid fijn stof als stikstofdioxide direct van invloed is op de totale sterfte en de sterfte door luchtwegaandoeningen en longkanker in Nederland. In de toekomst willen beide organisaties de gebruikte methodiek inzetten om ook de effecten van andere omgevingsfactoren op de gezondheid te onderzoeken.

Meer informatie: http://www.rivm.nl/Documenten_en_publicaties/Algemeen_Actueel/Nieuwsberichten/2015/Big_Data_geven_meer_kennis_over_gezondheidseffecten_luchtverontreiniging

4.

Veel kansen, veel uitdagingen



4.

Big data: Veel kansen, veel uitdagingen

Getuige de voorbeelden die hiervoor besproken worden, kan het beter benutten van de hoeveelheden data in en rond de gezondheidszorg van grote betekenis zijn. Wat is er nodig om deze ontwikkeling te versnellen? We bespreken in vogelvlucht een aantal uitdagingen die we moeten aangaan om de potentie van big data beter te benutten.

4.1 Technologie

Voor het verzamelen, opslaan en analyseren van big data sets zijn nieuwe technologieën nodig dan die waar bijvoorbeeld zorgaanbieders doorgaans over beschikken. De 'traditionele' architectuur is niet toereikend om de ontwikkelingen zoals in deze paper beschreven mogelijk te maken. Het is een ingewikkeld proces om het juiste technische fundament te plaatsen dat noodzakelijk is voor geavanceerde big data analyses. Het Amerikaanse National Institute of Standards and Technology benoemt als belangrijkste elementen daarvan de benodigde veranderingen in architectuur voor de opslag van data, het verwerken van data, het analyseren ervan en tot slot het visualiseren van de uitkomsten. Dit vereist, bijvoorbeeld van ziekenhuizen, een andere kijk op hun informatievoorziening.

Bij de voorbeelden in deze publicatie wordt meer dan eens data over de virtuele grenzen van de organisatie heen uitgewisseld. Meer integratie van databronnen vereist daardoor regelmatig ook nieuwe samenwerkingen tussen organisaties op verschillende niveaus. Hoe ziet deze samenwerking eruit en welke afspraken zijn daarvoor nodig? Welke data is

nodig en hoe wordt die geïntegreerd, wie gaat analyseren en hoe, wie gebruikt uiteindelijk de resultaten? Maakt men gebruik van technologieën waar de organisatie zelf over beschikt, open-source of niet, lokaal opgeslagen of in de cloud? Tal van vragen en uitdagingen waarover organisaties in de gezondheidszorg zicht de komende jaren zullen buigen.

4.2 Standaardisatie

De technologie wordt steeds geavanceerder en is steeds beter in staat om ongestructureerde informatie verwerken. Feit blijft dat het analyseren van gestructureerde data betrouwbaarder is. Een belangrijke uitdaging is dan ook om gegevens te structureren door de wijze waarop informatie wordt vastgelegd, uitgevraagd en uitgewisseld te standaardiseren. Hoewel er veel standaarden in de zorg worden toegepast, is interoperabiliteit tussen de informatiesilo's ver te zoeken (Nictiz, 2015). Zorgaanbieders registreren en verzamelen informatie op verschillende manieren in veel verschillende systemen. Uitwisseling daartussen is een moeizaam proces. Voor de groeiende hoeveelheid data die door patiënten wordt gegenereerd en opgeslagen, geldt hetzelfde. Door de wijze waarop gegevens worden vastgelegd in bronssystemen te standaardiseren, wordt de uitwisseling en de integratie ten behoeve van big data analyses eveneens makkelijker. Het is belangrijk dat publieke en private organisaties hiervoor (gezamenlijk) standaarden ontwikkelen en in gebruik nemen. Niet alleen op nationaal niveau, maar zeker ook internationaal. Het kan zijn dat we hier over tien of twintig jaar anders over denken, als verbeterde algoritmes in een handomdraai alle benodigde informatie uit natuurlijke (ongestructureerde) taal kunnen halen met dezelfde betrouwbaarheid als uit gestructureerde informatie. Echter, om vandaag en morgen al de potentiële waarde van alle gegevens te benutten, is standaardisatie van belang.

4.3 Toegang tot data

De fragmentatie van data tussen ziekenhuizen, patiënten, onderzoeksinstellingen, beleidsorganisaties, industrie en andere partijen is een belangrijke barrière voor het benutten van de potentie van big data. Elk van de datawarehouses van deze organisaties fungeert als een silo van potentiële informatie die lastig te openen is. Terwijl big data-analyses hun kracht juist ontlenen aan het integreren van de data uit deze verschillende silo's. Los

van de technische uitdaging die er ligt om verschillende databronnen te kunnen integreren, is het ook een uitdaging om eigenaren van data zo ver te krijgen dat ze deze überhaupt willen delen. In de gezondheidszorg heerst nog vaak een cultuur van “deze data zijn van mij”. Nieuwe samenwerkingen zijn nodig om dit patroon te doorbreken en aan te tonen hoe belangen elkaar kunnen aanvullen. Deze discussie krijgt sowieso een nieuwe dimensie nu patiënten steeds meer gegevens over zichzelf gaan beheren (Jacobs, 2015).

4.4 Privacy

Een andere discussie rondom het benutten van big data in de gezondheidszorg, gaat over de vragen wie eigenaar is van bepaalde vormen van gezondheidsinformatie, wie die informatie mag gebruiken en voor welke doeleinden. Er is wet- en regelgeving, zowel nationaal als Europees, over de bescherming van gezondheidsgegevens. Daarnaast zijn er gedragscodes onder beroepsgroepen en binnen instellingen die aanvullende eisen stellen aan de wijze waarop er omgegaan moet worden met privacygevoelige informatie. Absolute bescherming van privacy, in de zin van volledig controle hebben over de eigen informatie, lijkt onmogelijk. Onder meer omdat het in veel gevallen al niet eens duidelijk is wie de eigenaar is van de informatie of wie deze mag gebruiken. Mensen zijn zich er bijvoorbeeld vaak niet van bewust dat door

hun verzamelde gegevens ook gebruikt worden door commerciële partijen. Bovendien kunnen gebruikers van data zich zowel in het binnen- als het buitenland bevinden, waardoor ook wet- en regelgeving in andere landen relevant is.

Dit is ook een van de redenen waardoor gegevens die door burgers en betrokken zorgverleners worden verzameld, nog nauwelijks uitgewisseld worden. In een toekomstig scenario waarin mensen de zorg voor hun eigen gezondheid zo veel mogelijk vormgeven naar eigen wensen en behoeften, is breed gedragen, veilige en vertrouwde gegevensverwerking nóg crucialer dan dat dit nu het geval is, zo stelt de RVZ (2015). Er is een model nodig waarbij de privacy van patiënten geborgd wordt, zonder dat daarbij de noodzakelijke bewegingsvrijheid van gegevens tussen verschillende partijen in de gezondheidszorg in het geding komt. Nieuwe regelgeving vanuit de Europese Commissie zal daar ook een rol in spelen voor Nederland. Duiding van de wet- en regelgeving is nodig, zodat gebruikers van data weten wat de praktische consequenties ervan zijn. De RVZ waarschuwt ervoor dat indien dit niet gebeurt, de normen, kaders, faciliteiten en procedures door de verschillende commerciële partijen en hun verdienmodellen worden bepaald. Dit zou het benutten van de kansen van onder andere big data toepassingen voor burgers en maatschappij sterk kunnen beperken.

Registratie aan de Bron

‘Registratie aan de Bron’ is een meerjarenprogramma onder leiding van de Nederlandse Federatie van Universitair Medische Centra (NFU) en Nictiz. Het programma beoogt dat patiënten en zorgverleners in de toekomst altijd en overal over benodigde eenduidige gezondheidsinformatie beschikken. Het principe ‘registreren aan de bron’ betekent dat zorgverleners (en burgers) gezondheidsinformatie *eenduidig en eenmalig* vastleggen met behulp van de daarvoor ontwikkelde *zorginformatiebouwstenen*, volgens (inter) nationale standaarden. De vastgelegde gegevens kunnen daarna voor verschillende doeleinden worden gebruikt. Naast zorgverlening bijvoorbeeld ook voor overdracht tussen afdelingen en instellingen, kwaliteitsregistraties, wetenschappelijk onderzoek en administratieve processen.

De bouwstenen worden ontwikkeld door zorgprofessionals (artsen, verpleegkundigen en paramedici) zelf. Hierbij wordt zoveel mogelijk gebruik gemaakt van al bestaande (internationale) standaarden. Op dit moment zijn er circa 90 zorginformatiebouwstenen ontwikkeld. Daarbij kun je denken aan wilsverklaringen, medicatiegebruik, allergieën, lichaamsgewicht en familie anamneses.

Meer informatie: <https://www.nictiz.nl/projecten/specialistische-zorg/zorginformatiebouwstenen>

5. Tot slot

Deze publicatie belicht enkele van de vele inspirerende voorbeelden die laten zien hoe big data een positieve rol kan spelen in de gezondheidszorg. Toch is dit pas het topje van de ijsberg. De hoeveelheden data blijven toenemen en de mogelijkheden om er gebruik van te maken ook. Maar de praktijk is weerbarstig en we zijn de goudkoortsfase nog niet uit. Hopelijk lukt het de komende jaren om stap voor stap de uitdagingen het hoofd te bieden en het échte goud te delven: voorkomen dat mensen ziek worden en patiënten zo goed mogelijk helpen als dat toch gebeurt.

Geraadpleegde bronnen

- Alavi & Leidner (2001). Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues.
- Beyer & Laney (2012). The Importance of 'Big Data': A Definition. Gartner, Connecticut.
- Bruijn, De (2013). Naar een fraudebeeld Nederland. PwC, Amsterdam.
- Forbes (2012). Using Big Data to Help A Hospital Meet The Financial Future. Geraadpleegd op 11 november 2015. <http://www.forbes.com/sites/tomgroenfeldt/2012/04/20/aurora-health-uses-big-data-to-reduce-risk-in-outcomes-based-pay/>
- Grimes (2008). Unstructured Data and the 80 Percent Rule. Geraadpleegd op 11 november 2015. <http://breakthroughanalysis.com/2008/08/01/unstructured-data-and-the-80-percent-rule/>
- Groves, Kayyali, Knott & Van Keuken (2013). The Big Data Revolution in Healthcare. McKinsey&Company.
- Hengst, Van Pelt, Postema, Van der Sluis, De Jong, Van Dijk, Ekker & Pattel (2014). Zelfmetingen en de Nederlandse Gezondheidszorg. Nictiz, Den Haag.
- Huysman (2006). De zin en onzin van kennismanagement. Vrije Universiteit, Amsterdam.
- Institute for Health Technology Transformation (2013). Transforming Health Care Through Big Data.
- Jacobs (2015). Gebruik persoonlijke gezondheidsdata: nieuwe wetenschappelijke omgangsvormen. Geraadpleegd op 11 november 2015. <http://www.smarthealth.nl/trendition/2015/09/30/gebruik-persoonlijke-gezondheidsdata-nieuwe-wetenschappelijke-omgangsvormen/>
- Krijgsman & Ottenheijm (2015). Advies ter verbetering van informatie-uitwisseling in de zorg. Nictiz, Den Haag.
- Nederlandse Zorgautoriteit (2014). Onderzoek zorgfraude.
- Nationale Denktank (2014). Big Data in zicht.
- Raad voor de Volksgezondheid en Zorg (2015). Consumenten e-health. RVZ, Den Haag.
- Rathenau Instituut (2015). Big data en slimme algoritmen. Geraadpleegd op 11 november 2015. <https://www.rathenau.nl/nl/page/big-data-en-slimme-algoritmen>
- Soares (2012). A Framework that Focuses on the Data in Big Data Governance. Geraadpleegd op 11 november 2015. <http://www.ibmbigdatahub.com/blog/framework-focuses-data-big-data-governance>

