

Tussen computer en mens: betekenisgeving van medische data

Een onderzoek naar het waarde-creatieproces van medische data door ontwikkelaars en medici in een decision-support systeem

Masterscriptie

MA Bestuurs- en Organisationswetenschap

Publiek Management

Universiteit Utrecht

Begeleider Universiteit Utrecht

Marlot Kuiper MSc

Begeleider UMC Utrecht

dr. Teus Kappen

Tweede beoordelaar

dr. Robin Bouwman



Universiteit Utrecht



UMC Utrecht

Pien van Deth

3917894

Utrecht, 11 augustus 2019

Samenvatting

De medische wereld is aan verandering onderhevig door een toenemende complexiteit van informatie in de vorm van big data en aandacht voor *evidence-based* wetenschappelijke informatie (Murdoch & Detsky, 2013; Sim, 2016). Om de arts te helpen hier mee om te gaan in het besluitvormingsproces maken nieuwe technologieën als decision-support systemen met algoritmes een opmars in de medische wereld. In het UMC Utrecht wordt er gewerkt aan de ontwikkeling van een decision-support systeem genaamd UrSmartStatus. Deze applicatie zal de arts gaan ondersteunen in het besluitvormingsproces door informatie op een andere manier aan te bieden, een ernst-interpretatie te geven aan de actuele status van de patiënt en een prognose van de gezondheid van de patiënt te geven. Dit wordt gedaan door informatie in te delen per tractus in plaats van inputgericht en de organen groen, geel, oranje of rood te kleuren om de ernst van waardes aan te duiden. Dit wordt gedaan door een vergelijking te maken met *big data* van historische patiëntgevallen. De prognose zal daar ook op worden gebaseerd.

Met deze nieuwe vorm van informatie zal het besluitvormingsproces van de arts veranderen, omdat informatie op een andere manier wordt aangeboden én er een waarde-oordeel aan wordt gegeven. Iets wat de anders op dit moment zelf doet. In de app UrSmartStatus staat een rationele logica centraal door de algoritmische basis, maar er is wel sprake van ambiguïteit van medische data waar een rationele logica niet mee om kan gaan. Hoe kan de applicatie en dus de ontwikkelaars hier mee omgaan en hoe geven artsen hier hun interpretatie aan met hun rationele en klinische logica? De vraag is dus hoe hun besluitvormingsproces verandert met het gebruik van de app UrSmartStatus. In dit onderzoek stond dan ook de volgende onderzoeksvraag centraal: *Hoe krijgt medische data waarde in de decision-support applicatie UrSmartStatus in het UMC Utrecht?*

Uit de interviews met ontwikkelaars is naar voren gekomen dat hun waarde-creatie tweeledig is. Ten eerste de nieuwe manier van presenteren waarbij het werkproces van de arts meer centraal komt te staan. Ten tweede het waarde-oordeel wat aan de data wordt gegeven door middel van het kleuren van de organen aan de hand van het vergelijken van historische patiëntgevallen. Door het ambigu presenteren van de informatie zou er volgens de ontwikkelaars genoeg ruimte moeten overblijven voor de interpretaties van artsen. De kleuren vormen zodoende geen leidraad in het oordeel van de arts, maar nodigen uit om met een kritische blik tot een oordeel te komen. In een experiment is de interpretatie van de artsen getest door de artsen twee patiëntcasussen te geven waarvan de één verwerkt in het EPD en de ander in de applicatie, waarbij de mate van ambiguïteit van de casussen varieerde. De artsen werd gevraagd hun gedachteproces hardop uit te spreken tijdens het vormen van hun diagnose. Hieruit kwam voort dat er voor de artsen inderdaad genoeg ruimte overblijft voor eigen interpretatie. Hoewel de artsen benoemden de kleuren nuttig te vinden, gingen ze er niet (volledig) op af om een oordeel te vormen. Voor de ambigue casus leek de applicatie ondersteuning te bieden in het vergaren van inzicht over de patiënt, meer dan het EPD. Hieruit zou de voorzichtige conclusie kunnen worden getrokken dat de applicatie de arts inderdaad kan helpen in de rationele logica door dit ambigu te presenteren en aan te laten sluiten bij het werkproces van de arts, maar voldoende ruimte over te laten voor de menselijke, klinische logica die noodzakelijk is voor een medisch besluit.

Voorwoord

Beste lezer,

In dit afstudeeronderzoek komen de verschillende vakgebieden samen waar ik mij de afgelopen zeven jaar met interesse en plezier in heb verdiept. Hoe medisch professionals omgaan met innovaties in de vorm van decision-support systemen als bestuurs- en organisatievraagstuk. Welke implicaties het gebruik van algoritmes op een menselijk proces als medische besluitvorming kan hebben als ethisch-filosofisch vraagstuk. En hoe medische informatie zo goed mogelijk bij de arts terecht komt als communicatievraagstuk. Dit alles in de medische praktijk: een vakgebied waarmee ik onbekend was, maar een wereld waar ik het afgelopen half jaar mee kennis heb mogen maken. Een wereld die niet slaapt en waar een *overload* aan informatie dagelijkse praktijk is voor zorgverleners. Met bewondering voor hun omgang hiermee, heb ik mij ook in een overload aan informatie gestort en hier mijn eigen betekenis en waarde aan gegeven door mijn ervaring en intuïtie ermee te verbinden.

Een paar woorden van dank. Aan mijn scriptiebegeleider Marlot Kuiper, bij wie de deur altijd openstond voor vragen, advies en feedback op zowel inhoudelijk als procesmatig gebied. Op een fijne, persoonlijke en betrokken manier heeft zij mij geholpen en gestimuleerd een eigen draai te geven aan het onderzoek en daarin goede adviezen gegeven. Ze zorgde voor rust in het soms stressvolle proces. Aan Teus Kappen, wie ontzettend veel inspiratie en stof tot nadenken gaf. Hij stimuleerde mij en gaf de mogelijkheid om het onderzoek naar een hoger niveau te tillen. Met veel toewijding, energie, vertrouwen en af en toe een steuntje in de rug werd ik openhartig en op de beste manier begeleid. Aan Erik Koomen, die mij de mogelijkheid gaf een ‘kijkje in de keuken’ van de arts te krijgen; een indrukwekkende ervaring. Hij begeleidde mij in de complexiteit van medische besluitvorming en nam me mee in het hoofd van de arts. Met enthousiasme en gedrevenheid wordt het project UrSmartStatus door Teus en Erik geleid. Met dankbaarheid kijk ik terug op de fijne en leerzame samenwerking.

Het sluitstuk van een mooie periode ligt voor u. Veel leesplezier.

Pien van Deth

Utrecht, 11 augustus 2019

Inhoudsopgave

1. Inleiding	6
<i>Doel van het onderzoek</i>	8
<i>Onderzoeksvraag</i>	9
<i>Relevantie</i>	10
2. Theoretisch kader	13
2.1 <i>Decision-support systemen en voorspelmodellen</i>	13
2.2 <i>Algoritmes en het ethisch debat</i>	14
2.3 <i>Medische data: waarde-creatie</i>	16
2.4 <i>Medische data: ambiguïteit en onzekerheid</i>	16
2.4.1 <i>Ambiguïteit van medische data</i>	17
2.4.2 <i>Ambiguïteit van presentatie</i>	18
2.4.3 <i>Onzekerheid van medische data</i>	19
2.5 <i>Besluitvorming in de medische praktijk</i>	20
2.5.1 <i>Medisch oordelen en redeneren</i>	20
2.5.2 <i>Besluitvorming</i>	20
2.5.3 <i>Rationele besluitvorming</i>	21
2.5.4 <i>Medische besluitvorming</i>	22
2.6 <i>Conclusie en verwachtingen</i>	22
3. Methode	25
3.1 <i>Case study</i>	25
3.2 <i>Verkennde fase</i>	26
3.3 <i>Dataverzameling ontwikkelaars</i>	26
3.4 <i>Data-analyse ontwikkelaars</i>	27
3.5 <i>Data verzameling medici: experiment</i>	27
3.5.1 <i>Patiëntcasussen in EPD en app</i>	27
3.5.2 <i>Think-aloud methode</i>	28
3.5.3 <i>Experiment operationalisatie</i>	29
3.5.4 <i>Omschrijving van de setting</i>	29
3.5.5 <i>Hypotheses</i>	29
3.6 <i>Data-analyse medici</i>	30
3.7 <i>Betrouwbaarheid en validiteit</i>	32
3.7.1 <i>Betrouwbaarheid</i>	32
3.7.2 <i>Validiteit</i>	32
3.7.2.1 <i>Validiteit van patiëntcasussen</i>	32
3.7.2.2 <i>Validiteit van dataverzameling ontwikkelaars</i>	32
3.7.2.3 <i>Validiteit van dataverzameling medici</i>	32
4. Resultaten ontwikkelaars	36
4.1 <i>Verschillen tussen de applicatie en het EPD</i>	36
4.1.1 <i>Het doel van de applicatie</i>	37
4.1.2 <i>Keuzes in de applicatie</i>	38
4.2 <i>Besluitvorming: het verschil tussen de logica van de arts en de applicatie</i>	39
4.2.1 <i>Communicatie tussen de arts en de applicatie</i>	41
4.2.2 <i>Inzichtelijkheid van data en analyse</i>	43
4.2.3 <i>Vertrouwen van de arts in de applicatie</i>	43
4.3 <i>Interpretatie van data</i>	44
4.3.1 <i>Bias door missende data of artefacten</i>	46
4.3.2 <i>Ambiguïteit van data</i>	47
4.3.3 <i>Onzekerheid van data</i>	48
4.4 <i>Samenvattend</i>	48
5. Resultaten artsen	51

<i>5.1 Beschrijving experiment en analyse</i>	51
<i>5.2 Verschil in hoeveelheid geobserveerde informatie tussen app en EPD</i>	52
<i>5.3 Verschillen in besluitvormingsproces</i>	53
<i>5.4 Verschillen in differentiaal diagnose</i>	62
<i>5.5 Verschillen tussen de casussen met betrekking tot ambiguïteit</i>	63
<i>5.6 Gebruik van de applicatie</i>	64
<i>5.7 Samenvattend</i>	65
6. Conclusie	68
7. Discussie	71
8. Literatuur	73
Bijlagen	77

1. Inleiding

Algoritmische voorspelmodellen in de medische wereld

“The time approaches when algorithms, rather than oncologists, may take a lead role in deciding which combination of therapies is most likely to benefit a cancer patient. Computers not only will calculate the appropriate dose for radiation treatment (a responsibility that humans have already abdicated) but also may be pivotal in determining whether to use radiation at all.” (Senger & O’Leary in Cohen et al., 2018, pp.283-284).

In de medische wereld doet zich een toenemende complexiteit van informatie voor in de vorm van big data en ligt er meer focus op *evidence-based* wetenschappelijke informatie (Murdoch & Detsky, 2013; Sim, 2016). Hierdoor krijgt technologie in de vorm van (voorspel)algoritmes steeds meer aandacht. Aan de hand van big data worden deze algoritmes onder andere ingezet om medici te ondersteunen in hun besluitvorming (Senger & O’Leary in Cohen et al., 2018, pp.284). Een zogenaamd *decision-support* systeem is een computersysteem dat de besluitvorming van medici ondersteunt op het moment dat er besluiten moeten worden gemaakt bij individuele patiënten (Berner & La Lande, 2007 in Berner, 2007, p.3). Het is een hulpmiddel voor een medicus om een besluit te nemen, maar niet zomaar een hulpmiddel. Het is anders dan andere hulpmiddelen zoals een MRI-scan of een scalpel. Waar hulpmiddelen zoals deze laatste twee gezien kunnen worden als een vergroting van de menselijke capaciteit (het doet iets wat de menselijke capaciteit niet kan of minder goed kan), neemt een decision-support systeem een deel van de menselijke beoordeling over (pp.284-285). Waar de medicus de uitkomsten van een MRI-scan interpreteert en die informatie combineert met wetenschappelijke kennis en ervaring, kan een decision-support systeem al een deel van de interpretatie overnemen (Berner & La Lande, 2007 in Berner, 2007, p.4).

Decision-support systemen in groter ethisch mens-machine debat

Zulke decision-support systemen werken aan de hand van algoritmes. Een algoritme is een wiskundige constructie met een eindige reeks instructies of stappenplan om een bepaald doel te bereiken met de gegeven bepalingen (Mittelstadt et al., 2016, p.2). Deze zijn nooit waardevrij, omdat ze worden geprogrammeerd door mensen (Mittelstadt et al., 2016). Hierdoor moet er bij de implementatie ook zorgvuldig worden gekeken naar veiligheids- en ethische bezwaren (Senger & O’Leary in Cohen et al., 2018, pp.284). Het besluitvormingsproces van een medicus is namelijk moeilijk inzichtelijk te maken, doordat het deels gebaseerd is op ervaring en intuïtie, een universele logische structuur ontbreekt en het proces lastig te onderzoeken is zonder het proces zelf te beïnvloeden. Elke arts kan andere interpretaties maken op basis van dezelfde wetenschappelijke informatie. Toch wordt met een decision-support systeem een deel van het besluitvormingsproces overgenomen, waarin een algoritme interpretaties gaat maken en dus wel een bepaalde vastgezette logica zal volgen. Een technologie als decision-support staat daarmee in een groter mens-machine debat:

“If humans have had to accept the fact that machines drill better holes, paint straighter lines, have better memory... well, that is just the way the world is. But to suggest that machines can think better or more efficiently or to greater effect is to issue an extraordinary challenge” (Miller & Goodman in Goodman, 2009, p.102).

Het idee van een klinisch decision-support systeem is dat het besluitvormingsproces van een medicus wordt ondersteund. Daarmee wordt een deel van zijn of haar eigen proces overgenomen door het systeem, waardoor een gedeelte van het besluitvormingsproces wordt overgelaten aan het algoritme.

Een decision-support systeem intervenueert dus met het menselijk handelen en denken tijdens het besluitvormingsproces. Aan de ene kant kan dit gebeuren doordat een deel van de interpretatie over

wordt genomen, waardoor het waarde toekent aan de status of data van de patiënt (een deel van de diagnose). Aan de andere kant kunnen sommige decision-support systemen ook een voorspelling doen over het verloop van de gezondheid van een patiënt op basis van een menselijke database, individuele data en statistiek (Steyerberg, 2009, pp.1-2). De consequenties van beslissingen in de medische context gaan over de gezondheid en de levens van patiënten, waarin het draait om de beste behandeling voor de individuele patiënt. Hierdoor vraagt de implementatie van decision-support systemen om grote zorgvuldigheid en voorzichtigheid. Is het mogelijk, verantwoord en wenselijk om machines als decision-support systemen de besluitvorming van medici gedeeltelijk over te laten nemen?

Decision-support systeem UrSmartStatus geeft zowel actuele als prognostische status patiënt

In het Universitair Medisch Centrum Utrecht (hierna: UMC Utrecht) is er een nieuwe applicatie in ontwikkeling genaamd UrSmartStatus. De app is een vorm van een decision-support systeem. Het gaat een actuele visuele weergave geven van de gezondheid van de patiënt aan de hand van organen die rood, oranje of groen kleuren. Daarnaast zal het een prognose geven van het verwachte beloop van de gezondheid van patiënten die in het ziekenhuis zijn opgenomen. Het doel is door middel van de gebruiksvriendelijke en intuïtieve interface de zorgverleners het gemakkelijker te maken zich medische informatie eigen te maken, waardoor zij mogelijk sneller besluiten kunnen maken en wellicht in een later stadium ook betere beslissingen. De applicatie vertelt het verhaal van de patiënt visueel aan de hand van historische patiëntdata en algoritmes. De app zal complementair zijn aan het elektronisch patiëntendossier (hierna: EPD); het systeem wat de artsen op dit moment gebruiken. De precieze uitwerking en bepaling is nog in ontwikkeling.

Medische besluitvorming in verandering

De app in het UMC Utrecht kent dus twee veranderingen voor de dagelijkse praktijk van de medicus: de actuele informatie over de gezondheid van de patiënt wordt op een andere manier gepresenteerd waarmee de stand van de gezondheid ook al een waarde toegeschreven krijgt (rood, oranje, geel, groen). Daarnaast wordt er een prognose gegeven over de ontwikkeling van de gezondheid aan de hand van historische patiëntdata, individuele data en probabiliteit. Het grote verschil met het gebruik van het EPD is dat de medicus met het EPD zelf een soort mentale gezondheidsstatus van de patiënt moet maken uit alle losse, ‘feitelijke’ informatie en daar zelf een vertaalslag uit moet maken voor de actuele gezondheidsstatus en een prognose moet maken voor het verdere verloop van de gezondheid (Steyerberg, 2009, p.1). De app daarentegen geeft deze actuele gezondheidsstatus en prognose al en het is aan de arts om vervolgens de data erachter gedetailleerder te bekijken. Het algoritme in de app maakt gebruik van een rationele logica en statistische informatie (probabiliteit), terwijl een arts naast rationele logica ook zijn of haar klinische ervaring en intuïtie gebruikt (Peterson, 2009; Wang & Ruhe, 2007; Wilensky, 1964). Er zit dus een verschil in de manier waarop het algoritme data verwerkt en de manier waarop een medicus data verwerkt (zie ook paragraaf 2.5).

Prognoses van groot belang bij medische besluitvorming

Het verschil in informatieverwerking kan grote effecten hebben op het handelen en de besluitvorming van medici. Het stellen van prognoses is bijvoorbeeld één van de belangrijkste dingen in het medisch veld: “Prognosis is central to medicine. All diagnostic and therapeutic actions aim to improve prognosis” (Steyerberg, 2009, p.1). Medici maken besluiten en voorspellingen op basis van de prognoses, bijvoorbeeld welke behandeling of medicijnen er worden gekozen of de kans dat een bepaalde ziekte het verloop van de gezondheid van de patiënt veroorzaakt (pp.1-2). Met decision-support systemen kunnen de besluiten en voorspellingen een nog meer *evidence-based* karakter krijgen, doordat geschatte waarschijnlijkheden van risico’s en voordelen beter toegespitst kunnen worden op kenmerken van de individuele patiënt (p.2). Op basis van statistiek kan de waarschijnlijkheid van de prognose in kaart

worden gebracht (p.3). Dit is een verandering ten opzichte van de huidige besluitvorming: “[...] slechts een klein deel van ons medisch handelen [is] gebaseerd op bewijs” (Brand, 2012, p.1). Het onderzoek waarin de basis ligt voor dat bewijs is ook nog eens gevoelig voor vertekening en lang niet altijd toepasbaar op de individuele patiënt (Brand, 2012, p.1). Daarnaast zijn de overwegingen die medici maken voor de weging van het bewijs ook nog eens onduidelijk, waardoor ze tot verschillende interpretaties kunnen komen met dezelfde ‘evidence’ (Brand, 2012, p.1). Dit maakt het besluitvormingsproces van medici ook menselijk, omdat het vraagt om subjectieve of ervaringskennis om beslissingen te maken. Wetenschappelijke, evidence-based informatie is dus nooit eenduidig te gebruiken. Interpretatie heeft op dit moment een groot aandeel in de besluitvorming van medici, maar hoe vormt zich dit met een decision-support app?

Waarde-creatie in besluitvormingsproces

Interpretatie is met het gebruik van een decision-support app dus niet meer alleen het werk van de medicus, maar ook van de ontwikkelaar van de app (en het algoritme). Voor de medicus betekent dit dat hij of zij op een andere manier waarde toekent aan de data uit de app door die te interpreteren, maar ook dat waarde wordt gecreëerd door de ontwikkelaars. Waarde-creatie in de vorm van interpretatie is dus niet meer alleen een proces van de medicus, maar ook van de ontwikkelaars. Beide geven betekenis aan de medische data. Er wordt van data betekenisvolle informatie gemaakt door zowel de ontwikkelaars als de artsen.

Onzekerheid en ambiguïteit in medische data

De waarde-creatie van medische data is echter lastig. Statistische modellen waar het algoritme mee werkt, zijn niet zo gemakkelijk vorm te geven als het doet lijken. Voor het herkennen van patronen in data moeten assumpties worden gemaakt (Steyerberg, 2009, p.3), terwijl er in medische data veel onzekerheid en ambiguïteit zit (Kappen, 2015, pp.108-109). Onzekerheid speelt altijd een rol omdat een diagnose of prognose nooit met 100% zekerheid of waarschijnlijkheid kan worden gegeven. Ambiguïteit van de data komt voort uit de interpretatie die aan de data moet worden gegeven. Dezelfde gegevens kunnen namelijk anders geïnterpreteerd worden wanneer iemand een hartpatiënt of een longpatiënt is, wat gevolgen heeft voor de diagnose die wordt gesteld en de behandeling die wordt gekozen. De vraag is hoe ontwikkelaars omgaan met deze onzekerheid en ambiguïteit van medische data bij het ontwikkelen van (algoritmes in) de app, hoe zij hier inhoud aan geven en hoe dit wordt geïnterpreteerd door medici. Ambiguïteit en onzekerheid van data zijn namelijk bepalend voor de interpretatie van medische data en daarmee voor de medische besluitvorming. De medische data moet waarde krijgen, maar hoe en wanneer wordt dit gedaan en is dit inzichtelijk?

Doel van het onderzoek

In dit onderzoek wil ik inzicht creëren in het proces van waarde-creatie van medische data in algoritmes van de app UrSmartStatus. Door het gebruiken van de decision-support app UrSmartStatus zal het waarde-creatie proces van medische data veranderen. Waar medici nu de enige zijn die betekenis geven aan de data uit het EPD, zullen ontwikkelaars met de implementatie van de app ook een rol spelen in het proces. Medici krijgen met de app waarde-geladen informatie aangeboden en maken op basis daarvan hun besluit. Dit kan de informatieverwerking van medici veranderen, wat invloed kan hebben op hun werkproces en uiteindelijk wellicht ook hun besluiten en verantwoordelijkheid voor dat besluit. Met dit onderzoek wil ik erachter komen hoe de waarde door de app-ontwikkelaars gecreëerd worden en hoe medici de data interpreteren. Is de waarde-creatie van de ontwikkelaars (en daarmee het algoritme) inzichtelijk, duidelijk en transparant en is dit nodig voor de interpretatie van de medici? Algoritmes zijn waarde-geladen en veranderen daarmee het proces. Welke waarde wordt er toegevoegd en op welke manier beïnvloedt de betekenisgeving van de ontwikkelaars de interpretatie van de artsen?

Kortom, kan de decision-support app UrSmartStatus de medici inderdaad ondersteuning bieden bij de interpretatie van medische data met betrekking tot ambiguïteit en onzekerheid?

Het evalueren van het besluitvormingsproces van medici met het gebruik van de app UrSmartStatus kan daarnaast voor de ontwikkelaars belangrijke inzichten geven. De app is ontwikkeld om de medicus te ondersteunen in het besluitvormingsproces, dus inzicht in dat besluitvormingsproces is van belang voor het maken van een optimaal decision-support systeem (Todd & Benbasat, 1987, p.501). Op die manier kunnen problemen die zich voordoen tijdens het besluitvormingsproces aan het licht komen (p.501). Daarnaast kan het experimentele gedeelte van dit onderzoek een blauwdruk vormen voor de manier waarop de applicatie wordt getest. Het experiment kan namelijk op zichzelf ook gezien worden als een onderzoek naar hoe applicaties en andere systemen die implicaties hebben voor het besluitvormingsproces en werkproces van artsen getest kunnen worden.

In de dagelijkse praktijk van de medicus is ook de patiënt onderdeel van de besluitvorming met wensen en voorkeuren en is het dan vaak ook een gezamenlijk besluitvormingsproces. In dit onderzoek zal dit echter niet worden meegenomen, omdat het specifiek gaat om de ondersteuning van het besluitvormingsproces van de medicus en het gebruik van de app.

Onderzoeksvraag

Om het proces van waarde-creatie van ontwikkelaar tot arts in kaart te brengen is de volgende onderzoeksvraag opgesteld:

Hoe krijgt medische data waarde in de decision-support applicatie UrSmartStatus in het UMC Utrecht?

Om deze vraag te beantwoorden zal er naar twee processen van waarde-creatie worden gekeken, namelijk de ontwikkeling van de app en het gebruik van de app. Aan de ene kant gaat het om de waarde die ontwikkelaars toekennen aan medische data en hoe zij dit verwerken in de app, aan de andere kant de waarde die medici toekennen aan de informatie uit de app en dit interpreteren. Bij het waarde-creatie proces zal specifiek worden ingegaan op ambiguïteit en onzekerheid van medische data. Om dit te operationaliseren, zullen er interviews worden gehouden bij de ontwikkelaars en zal het interpretatieproces van de artsen aan de hand van een experiment worden geduïd.

Deelvragen

De hoofdvraag wordt getracht te beantwoorden aan de hand van theoretische en empirische deelvragen.

Theoretische deelvragen:

- Wat zijn decision-support systemen en welke uitdagingen brengen zij met zich mee?
- Wat zijn algoritmes en wat zijn eventuele implicaties voor de toepassing ervan?
- Wat wordt bedoeld met waarde-creatie van medische data?
- Wat betekent de ambiguïteit en onzekerheid van medische data?
- Hoe werkt rationele en medische besluitvorming?

Empirische deelvragen:

- Hoe wordt waarde toegekend aan medische data door de ontwikkelaars van UrSmartStatus?
- Hoe wordt waarde toegekend aan informatie uit de applicatie UrSmartStatus door medici?

Om de empirische deelvragen te operationaliseren, is bij de ontwikkelaars specifiek gevraagd naar ambiguïteit en onzekerheid van data en daarbij gevraagd naar de uitwerking van de actuele status en de

prognose van de gezondheid van de patiënt. Daarnaast is ingegaan op hoe zij het verschil en de communicatie zien tussen de logica van de app en de algoritmes en de logica van artsen. Bij de medici is enkel gefocust op ambiguïteit van medische data en dus de actuele status van de patiënt, omwille van de haalbaarheid en correctheid van het experiment. Er is dus niet ingegaan op hoe zij de prognose zouden interpreteren van de gezondheid van patiënten.

Relevantie

Het waarde-creatie proces van medische data verandert met de komst van een decision-support systeem als de applicatie UrSmartStatus. Waar artsen nu zelf een interpretatie maken van de data, gaan straks ook ontwikkelaars betekenis geven aan die data waardoor de artsen de data op een andere manier krijgen gepresenteerd. Dit onderzoek kan een bijdrage leveren aan de ontwikkeling van de applicatie, inzicht in (de verandering van) het proces van betekenisgeving en interpretatie van medische data met tussenkomst van algoritmes en inzicht in de communicatie tussen arts en algoritme.

Voor ontwikkelaars van decision-support systemen is het van belang om inzicht te hebben in het besluitvormingsproces van de gebruikers (Todd & Benbasat, 1987, p.501). Het zijn interactieve systemen waarbij de gebruiker afhankelijk is van het design van in dit geval de app voor het begrip van de informatie (p.501). Om de werking en effectiviteit van de app te meten, moet er gekeken worden naar hoe problemen worden opgelost en besluitvorming plaatsvindt door gebruikers, de medici. Verandering in de manier van probleemoplossing en besluitvorming kan iets zeggen over de effectiviteit en werking van een decision-support systeem (pp.501-502): “the evaluation of a DSS [decision-support systeem] is thus based not only on simple output measures but also on understanding the changes in the process of decision making” (p.502). Daarnaast dient het experiment als een experiment op zichzelf om te onderzoeken hoe op de best mogelijke manier in beeld kan worden gebracht wat de implicaties zijn van een decision-support systeem op het besluitvormingsproces van artsen.

Maatschappelijke relevantie

Wanneer samenwerking tussen ‘mens en machine’, in dit geval medicus en decision-support systeem, goed verloopt, kunnen beide elkaar versterken (Beerends, 2019). De zorg kan in dat geval op een positieve manier veranderen doordat processen sneller gaan en wellicht ook beter. Om die samenwerking goed te laten verlopen kan dit onderzoek een bijdrage leveren aan de ontwikkeling en eventuele implementatie van de applicatie. Het kan voor de ontwikkelaars inzicht bieden in hoe zij waarde toevoegen aan medische data in de app en in de interpretatie en besluitvormingsproces van de medisch gebruikers. Door Li et al. (2012) wordt ook gesteld dat CDSS vaak geïmplementeerd worden zonder dat er degelijke evaluatie onderzoeken zijn gedaan (p.769). Dit onderzoek kan tevens dienen als evaluatie-onderzoek voor de implementatie van de decision-support app UrSmartStatus. Het doen van gedegen *usability testing* is noodzakelijk voordat een dergelijk systeem wordt geïmplementeerd (Li et al., 2012, p.762). Dit onderzoek kan een eerste aanzet daartoe zijn.

Dit inzicht van ontwikkelaars is ook belangrijk voor de ethische implicaties die algoritmes met zich meebrengen. Algoritmes worden door mensen gemaakt en reflecteren daarmee ook hun waarden. De ontwikkelaars voegen waarde toe aan de betekenisgeving van de medische data, maar het is nog niet bekend hoe dit geïnterpreteerd zal worden door artsen en op welke manier dit de besluitvorming kan veranderen. Het verantwoordelijkheidsvraagstuk speelt een rol wanneer technologie wordt gebruikt, maar de logica niet volledig inzichtelijk is voor de gebruiker (Mittelstadt et al., 2016, pp.10-12). Zeker in een medische context is dit van belang, omdat besluiten implicaties kunnen hebben voor de gezondheid of het leven van patiënten. De vraag wie er verantwoordelijk is als een decision-support systeem is gebruikt, is lastig te beantwoorden.

Met het gebruik van technologie veranderen ook de standaarden binnen de zorg. Dit heeft effect voor de manier waarop medici hun werk inrichten. Als het gebruik van technologie op een gegeven moment standaard wordt of zelfs verplicht wordt, verandert de routine en de taakinvulling van de medicus (Goodman, 2007 in Berner, 2007, p.130). Het gebruiken van computers als standaard in de zorg vraagt om zorgvuldigheid. De technologie moet goed werken, transparant zijn en op de juiste manier gebruikt worden om het juiste effect te kunnen bereiken. Door inzicht te krijgen in hoe ontwikkelaars waarde geven aan data en medici interpretatie geven aan ambigue en onzekere data uit de applicatie kan inzicht verkregen worden in hoe een dergelijke app wel of niet zou kunnen passen in de zorg.

Wetenschappelijke relevantie

Dit onderzoek kan een bijdrage leveren aan de literatuur over klinische decision-support systemen door een experimentele opzet te geven om op zowel een kwalitatieve als kwantitatieve manier de effecten van decision-support systemen op het besluitvormingsproces van artsen in kaart te brengen. Inhoudelijk kan het inzicht geven in welk opzicht het besluitvormingsproces van de arts verandert en welke waarde ontwikkelaars kunnen toevoegen in het proces. Duidelijk zal worden hoe ontwikkelaars tegen het implementeren van een decision-support systeem aankijken en hun eigen rol daarin zien, en op welke manier artsen deze waarde-creatie interpreteren. In het onderzoek wordt er zowel gekeken naar de waarde-creatie van de ontwikkelaars middels interviews, als naar de interpretatie van de artsen middels een experiment met zowel kwalitatieve als kwantitatieve elementen. Zodoende wordt het gehele proces van waarde-creatie in beeld gebracht.

Methodologisch kan dit onderzoek een nieuwe toevoeging zijn aan de literatuur over de think-aloud techniek in een klinische omgeving. Bij deze techniek denken de participanten hardop na over de stappen die zij zetten bij het oplossen van een probleem of een besluitvormingsproces (Lundgrèn-Laine & Salanterä, 2010; Todd & Benbasat, 1987). Volgens Lundgrèn-Laine en Salanterä (2010) is er nog weinig geschreven over deze techniek specifiek in de klinische setting. Ook de combinatie met protocol analyse wordt niet vaak gemaakt (pp.565-566). Een protocol analyse houdt in dat de stappen die de participant zet in het gedachteproces in kaart worden gebracht (Lundgrèn-Laine & Salanterä, 2010; Todd & Benbasat, 1987). Ook Todd en Benbasat (1987) noemen dat er met betrekking tot onderzoek naar decision-support systemen, weinig aandacht is voor het begrijpen van het proces. Protocol analyse als process-tracing methode kan de impact van een decision-support systeem op het besluitvormingsproces in kaart brengen (p.506).

Todd en Benbasat (1987) geven aan dat protocol analyse ook door ontwikkelaars gebruikt kan worden tijdens de ontwikkeling van de app. Zij stellen dat er weinig verschil is tussen een protocol analyse om domein-specifieke kennis te vergaren en het testen van de kwaliteit van de *interface*. Beide manieren geven feedback op het gebruik van het systeem (p.507). Protocol analyse kan in dit onderzoek als toepassing dienen voor *usability testing* van decision-support systemen (Li et al., 2012). Het kan voor ontwikkelaars kennis bieden over hoe medici omgaan met de applicatie en over de begrijpelijkheid van zowel de informatie als het design (Todd en Benbasat, 1987, p.506).

Leeswijzer

Voor het beantwoorden van de hoofdvraag zal eerst in hoofdstuk 2 de theoretische deelvragen worden beantwoord. In dit theoretisch kader zal eerst de literatuur worden besproken over decision-support systemen. Daarna zal de werking van algoritmes als essentieel onderdeel van decision-support systemen en de ethische implicaties van algoritmes met betrekking tot dit onderzoek worden uiteengezet. Vervolgens wordt uitgelegd wat er wordt verstaan onder waarde-creatie van medische data en de rol van

ambigüiteit en onzekerheid hierin. De medische data is de basis voor de medische besluitvorming. Er zal dan ook worden verklaard hoe dit proces van besluitvorming werkt, met een onderscheid tussen de theorie over rationele besluitvorming, waarop algoritmes zijn gebaseerd, en het intuïtieve en reflectieve component van medische besluitvorming. In hoofdstuk 3 zullen de methodes beschreven worden voor de dataverzameling en –analyse, interviews voor de ontwikkelaars en een experiment voor de artsen, waarmee de empirische deelvragen worden getracht te beantwoorden. In hoofdstuk 4 en 5 worden de resultaten en bevindingen hiervan besproken en wordt ook stilgestaan bij de vergelijking van waardecreatie tussen de ontwikkelaars en artsen. In hoofdstuk 6 zal een conclusie worden getrokken en de hoofdvraag worden beantwoord. Ten slotte worden er in hoofdstuk 7 de discussiepunten van dit onderzoek besproken en aanbevelingen gedaan voor vervolg- of vergelijkbaar onderzoek.

2. Theoretisch kader

Om de hoofdvraag te beantwoorden zal ik eerst ingaan op literatuur die inzicht geeft in het proces van waarde-creatie in een decision-support app. Hierin zullen de theoretische deelvragen worden beantwoord. Ten eerste zal het fenomeen decision-support systemen nader worden toegelicht met de belangrijkste knelpunten die zich daarbij voordoen (sectie 2.1). Vervolgens wordt stilgestaan bij de werking van algoritmes en het ethisch debat eromheen (sectie 2.2). Een algoritme krijgt input binnen en geeft een output op basis van een bepaalde logica. De input is in dit geval medische data en de output is de actuele status van een patiënt en een prognose van zijn of haar gezondheid. De medische data krijgt dus ergens tussen input en output waarde. Wat er wordt bedoeld met het waarde-creatie proces wordt toegelicht in sectie 2.3. Dit proces kent ook ambiguïteit en onzekerheid, wat de waardebepaling lastig maakt. De ambiguïteit en onzekerheid zal nader worden toegelicht in sectie 2.4. Op basis van de informatie uit het decision-support systeem maakt de medicus een besluit. Hoe deze besluitvorming werkt en wat de verschillen zijn tussen de besluitvorming van algoritmes en mensen wordt besproken in sectie 2.5. Ten slotte wordt in sectie 2.6 een conclusie getrokken uit de theorie en verwachtingen geformuleerd ten aanzien van het empirisch onderzoek.

2.1 Decision-support systemen en voorspelmodellen

In deze sectie zal de volgende theoretische deelvraag worden beantwoord: *wat zijn decision-support systemen en welke uitdagingen brengen zij met zich mee?* Voor het empirisch onderzoek is het van belang om te weten wat decision-support systemen precies kunnen doen en welke uitdagingen er bekend zijn uit de literatuur voor zowel de ontwikkelaars als de gebruikers. Dit kan namelijk inzicht geven in en verwachtingen geven over hoe een decision-support systeem werkt en welke punten lastig kunnen zijn voor ontwikkelaars en gebruikers.

Steeds meer gezondheidszorgorganisaties wereldwijd implementeren decision-support systemen om de efficiëntie en veiligheid van de zorg te verbeteren (Li et al., 2012, p.762). Klinische decision-support systemen zijn “computer systems designed to impact clinician decision making about individual patients at the point in time that these decisions are made” (Berner & La Lande, 2007 in Berner, 2007, p.3). Het design houdt in dat er een database is, een modelbase (het model waarin de data verwerkt wordt) en een interface voor de interactie tussen de gebruiker en het computersysteem waarin de verwerkte data wordt gepresenteerd (Todd & Benbasat, 1987, p.501). In deze systemen kunnen ook voorspelmodellen een rol spelen, zoals ook zal gebeuren in de decision-support app UrSmartStatus in het UMC Utrecht. Klinische voorspelmodellen kunnen omschreven worden als voorspellingsregels of prognostische modellen die een aantal kenmerken of variabelen, gerelateerd aan de patiënt, de ziekte of de behandeling, combineert om een diagnostische of prognostische uitkomst te voorspellen (Steyerberg, 2009, p.2).

Clinical decision-support systems (CDSS) helpen medici besluiten te maken waarmee wordt getracht de zorg efficiënter en effectiever te maken (Li et al., 2012, p.762). Zo zouden de systemen kunnen helpen om medici sneller besluiten te kunnen laten maken waardoor er tijd wordt bespaard. Daarnaast zou het er ook voor kunnen zorgen dat medici betere beslissingen maken doordat ze meer informatie hebben en/of informatie beter gecommuniceerd en gepresenteerd kan worden. Veel CDSS zijn echter nog niet zo ver dat dit ook realiteit is (p.762). Hoewel het gedrag van medici vaak positief verandert door het effect van herinneringen, alarmen, behandelplannen en patiënteducatie (Kaplan, 2001, p.17), komt er ook uit veel onderzoeken naar voren dat het gebruik en de effecten van CDSS nog verbeterpunten kent (o.a. Kaplan, 2001; Sim et al., 2001). Een aantal van deze punten zijn dat medici niet altijd welwillig tegenover het gebruik van CDSS staan (o.a. Esmaeilzadeh, Sambasivan, Kumar & Nezakati, 2015; Kappen, 2015), de impact van CDSS op de patiëntresultaten niet altijd duidelijk zijn (Cresswell, Majeed, Bates, & Sheikh, 2012; Garg et al., 2005; Hunt, Haynes, Hanna, & Smith, 1998), de CDSS niet altijd

geschikt zijn voor individuele patiënten (met co-morbiditeit) of voor specifiek klinische omstandigheden (Légaré, Ratté, Gravel, & Graham, 2008; Sittig et al., 2008) en de CDSS nog vaak niet passend geïntegreerd kunnen worden in de workflow van medici (Cresswell et al., 2012; Kawamoto, Houlihan, Balas, & Lobach, 2015; Li et al., 2012, p.762).

CDSS zijn in hun opmars dus nog niet eenduidig in hun werking en mogelijke voor- en nadelen. Ditzelfde zal gelden voor de ontwikkeling van een decision-support systeem als UrSmartStatus. In dit onderzoek zal ik bekijken hoe het ontwikkelingsproces vanuit de ontwikkelaars eruitziet, hoe zij waarde toevoegen en op welke manier artsen dit interpreteren. Door op deze specifieke casus in te zoomen is het mogelijk om de voor- en nadelen van dit soort systemen bloot te leggen, zowel vanuit een ethisch-maatschappelijk als technologisch perspectief.

2.2 Algoritmes en het ethisch debat

CDSS zijn systemen die grote hoeveelheden stukjes data verwerken. De verwerking van deze data wordt in de applicatie door algoritmes gedaan. In deze sectie zal antwoord worden gegeven op de deelvraag *wat zijn algoritmes en wat zijn eventuele implicaties voor de toepassing ervan?* De decision-support app UrSmartStatus in het UMC Utrecht werkt met algoritmes. Om meer inzicht te krijgen in de factoren waar de ontwikkelaars rekening mee moeten houden, zal het artikel ‘The Ethics of Algorithms: Mapping the Debate’ van Brent Mittelstadt et al. (2016) worden besproken. Hierin wordt een overzicht gegeven van het debat rondom de ethiek van algoritmes. Het artikel van Mittelstadt et al. kan gezien worden als een overzichtsartikel van de wetenschappelijke literatuur over de ethische aspecten van algoritmes.

Een algoritme kan omschreven worden als “a mathematical construct with a finite, abstract, effective, compound control structure, imperatively given, accomplishing a given purpose under given provisions” (Hill, 2015 in Mittelstadt et al., 2016, p.2). De auteurs voegen hier nog aan toe dat wanneer we het hebben over een algoritme, dit niet alleen gaat over de wiskundige constructie maar ook om de implementatie in een technologie (p.2), zoals in dit geval de applicatie in het UMC Utrecht. Die technologie wordt gebruikt voor een specifieke taak, dus het algoritme wordt hier op aangepast (p.2).

Door de wiskundige basis lijkt het vaak dat deze algoritmes op een ‘objectieve’ manier data verzamelen en een output geven, waarbij een output bijvoorbeeld een voorgestelde route kan zijn. Dit wordt ook wel de ‘computationele denkfout’ genoemd. Dit is de gedachte dat wat een computer als uitkomst geeft meer valide, nauwkeurig of betrouwbaar is dan wat een mens kan creëren (Goodman in Berner, 2007, p.132). Het automatiseren van menselijke besluitvorming wordt om die reden vaak ook gerechtvaardigd (Bozdog, 2013; Naik & Bhide, 2014 in Mittelstadt et al., 2016, p.7). De decision-support app UrSmartStatus kan ook gezien worden als een gedeeltelijke automatisering van de menselijke besluitvorming.

Algoritmes zijn echter altijd waarde-geladen (Mittelstadt et al., 2016, p.1,7). Ze worden gecreëerd door ontwikkelaars, in dit geval in samenwerking met artsen, die bewust of onbewust keuzes moeten maken en daarbij bepaalde waarden als belangrijker achten dan andere (p.1): “the values of the author, wittingly or not, are frozen into the code, effectively institutionalising those values” (Macnish, 2012, p.158). Het is lastig om de impact hiervan te overzien, omdat de invloed van de subjectiviteit van mensen lastig te achterhalen is (Mittelstadt et al., 2016, p.2). Voor de ontwikkelaars in het UMC Utrecht betekent dit dat zij gedeeltelijk betekenis of waarde geven aan de medische data, wat anders de medicus zou doen. Dit vraagt dus om zorgvuldigheid, omdat het een interventie is in het besluitvormingsproces van de medicus en de besluiten van levensbelang kunnen zijn. De waarde-creatie van de ontwikkelaars is dus geen

neutraal proces. Het is daarom interessant te bekijken op welke manier zij waarde creëren in het besluitvormingsproces en waarde geven aan de medische data.

Wanneer het besluitvorming betreft, verschilt de logica van algoritmes met de logica van mensen. Dit wordt uitgebreider uiteengezet in paragraaf 2.5. Algoritmes kunnen veel meer data verwerken die relevant zijn voor een beslissing (Mittelstadt et al., 2016, p.3). Zodoende kunnen er met grote hoeveelheden data complexe analyses uitgevoerd worden. Dit maakt de zorgvuldigheid waarmee algoritmes data verwerken nog belangrijker, omdat de impact ervan groot kan zijn. De applicatie UrSmartStatus heeft enkel ‘handgemaakte’ algoritmes, waardoor het in ieder geval voor de ontwikkelaars duidelijk is welke regels en welke waarden worden gehanteerd. Dit is anders bij *machine-learning* algoritmes. Deze algoritmes maken zelf regels gebaseerd op de aanwezige data waarmee nieuwe data beoordeeld wordt (p.3). Dat geeft het proces onzekerheid en ondoorzichtigheid, doordat het algoritme een ‘black-box’ heeft (p.3,6). Toch kunnen handgemaakte algoritmes ook zo complex zijn dat de logica voor de mens onnavolgbaar is (p.6). Het black-box probleem heeft te maken met de transparantie van algoritmes. Door de onnavolgbaarheid van algoritmes zijn ze soms moeilijk te controleren of corrigeren (Tutt, 2016 in Mittelstadt et al., 2016, p.6). De twee belangrijkste onderdelen van transparantie zijn de toegankelijkheid en begrijpelijkheid van informatie. Het is echter heel lastig om een complex besluitvormingsproces van een algoritme toegankelijk en begrijpelijk te maken (p.6), net als dat bij mensen lastig is. Een groot verschil tussen het besluitvormingsproces van een algoritme en van een mens is dat de logica van een algoritme vaak onnavolgbaar is, maar een mens kan zijn proces verwoorden (p.7). De legitimiteit van beslissingen is bij algoritmes dus soms lastig te achterhalen, terwijl mensen dit wel in zekere mate kunnen verwoorden (p.7). De beslissingen die anders door medici zouden worden gemaakt, en die in zekere mate voor hen zelf navolgbaar zijn, worden nu dus gedeeltelijk overgenomen door het algoritme. De ontwikkelaars maken hierin de beslissingen. Voor medici kan het dus belangrijk zijn om de transparantie van de keuzes van ontwikkelaars te hebben in de app, om te zien hoe bepaalde diagnoses of prognoses tot stand zijn gekomen. Zij moeten namelijk vertrouwen op een applicatie of de beslissingen van ontwikkelaars in het komen tot hun besluit, in plaats van dat zij zelf alle data verzamelen en categoriseren en op basis daarvan diagnoses en prognoses stellen.

Een ander lastig fenomeen bij algoritmes is de interpretatie van de output. Een onderdeel hiervan zijn de correlaties die gemaakt worden. Algoritmes halen uit grote datasets correlaties, waarop de gebruiker handelingen baseert maar waarvan de causaliteit niet is aangetoond en ook lastig aan te tonen is, omdat de correlaties vaak niet reproduceerbaar of falsifieerbaar zijn (Mittelstadt et al., 2016, p.5). Voor het verklaren van correlaties is rechtvaardiging nodig, statistische modellen hebben op zichzelf geen betekenis. Die betekenis moet eraan worden gegeven (p.8). Het geven van die betekenis aan of het handelen op basis van het algoritme kent echter ook problemen. Door de subjectiviteit van algoritmes en de filtering van informatie kan het gedrag van de gebruiker beïnvloed of zelfs genudged worden (p.9) Dit kan de autonomie van de gebruiker aantasten. Het lastige hierbij is dat het filteren van informatie ook juist gebruikt wordt om de besluitvorming van de gebruiker te vergemakkelijken door alleen de relevante informatie aan te bieden (p.9).

De ethische implicatie van algoritmes is belangrijk om in acht te houden, zeker gezien de context van medische besluitvorming. Afwegingen die medici maken, hebben gevolgen voor de gezondheid of soms zelfs het leven van mensen, waardoor ze zwaar wegen. Het implementeren en gebruiken van decision-support systemen vereist dus grote zorgvuldigheid. Daarnaast speelt mee dat het besluitvormingsproces van medici, of überhaupt mensen in het algemeen, geen inzichtelijk, transparant proces is. (Medisch) professionals gebruiken hun ervaring en impliciete kennis om tot besluiten te komen (Mittelstadt et al., 2016, p.11). Met tussenkomst van technologie is het belangrijk om hier rekening mee te houden, omdat

verandering in zowel het fysieke als cognitieve proces consequenties kan hebben voor de besluiten, maar door het niet-transparante cognitieve proces zullen deze consequenties niet direct navolgbaar zijn: “When algorithmic and human decisionmakers work in tandem, norms are required to prescribe when and how human intervention is required, particularly in cases like high-frequency trading where real-time intervention is impossible before harms occur” (Mittelstadt et al., 2016, p.11).

2.3 Medische data: waarde-creatie

Ter verduidelijking van het begrip ‘waarde-creatie’ wordt er in deze paragraaf antwoord gegeven op de deelvraag: *wat wordt bedoeld met waarde-creatie van medische data?* Met het gebruiken van een decision-support applicatie als UrSmartStatus verandert het waarde-creatie proces van medische data. Met waarde-creatie wordt bedoeld het geven van betekenis aan data. In de huidige situatie met de huidige systemen die de artsen van het UMC Utrecht gebruiken betekent dit dat er een interpretatie wordt gegeven aan de medische data waarmee een beeld van de gezondheid van de patiënt kan worden geschetst. Het geven van betekenis of een interpretatie aan data houdt in dat een arts bepaalde data kan bestempelen met een waarde: de gegeven data kan in de context van de patiënt betekenen dat het goed of slecht gaat; er wordt een waarde aan gegeven. Wanneer bijvoorbeeld de bloeddruk van een patiënt wordt gemeten, is de uitslag daarvan gegeven in getallen (data). Op zichzelf zeggen die getallen niets, maar met de kennis over wanneer bloeddruk (te) hoog of (te) laag is en kennis over de patiënt (leeftijd, ziektebeeld) kan er betekenis of waarde worden gegeven aan de uitslag. Bepaalde getallen zullen betekenen dat de patiënt te hoge bloeddruk heeft, andere getallen zullen betekenen dat de patiënt te lage bloeddruk of een goede bloeddruk heeft. Waarde-creatie betekent in dit onderzoek dus niet waarden in een politiek discours zoals vrijheid of autonomie (Geuijen et al., 2017) en ook niet waardes als kwaliteiten, zoals betrouwbaarheid of efficiëntie (Van der Wal, De Graaf, & Lasthuizen, 2008). In dit onderzoek betekent het betekenis of interpretatie geven aan data. Wanneer gesproken wordt over het meervoud van het woord zal de term ‘waarde’ worden gebruikt. Voor het aanduiden van medische data/getallen zal gebruik worden gemaakt van de term ‘waardes’.

Het waarde-creatie proces verandert met het gebruiken van een decision-support systeem, omdat er in de applicatie al waarde wordt gegeven aan de medische data. De arts interpreteert daarmee niet alleen de getallen, maar ook de waarde die eraan is gegeven door het algoritme of de ontwikkelaars. Zoals in de vorige paragraaf ook beschreven, zijn algoritmes altijd waarde-geladen (Mittelstadt et al., 2016, p.1,7). Ontwikkelaars maken namelijk bedoeld of onbedoeld keuzes in het model en de toepassing. Medische data is een bijzondere vorm van data, omdat het op allerlei manieren *bias* kent; ontbrekende informatie. Daarbij zijn de variabelen en relaties van informatie niet geheel duidelijk (Kappen, 2015, p.108). De data is namelijk altijd contextafhankelijk en niet altijd volledig. Omdat een algoritme werkt volgens een mathematisch model, is deze bias problematisch. Een mathematisch model kan vanuit zichzelf namelijk geen *bias* herkennen, daar is menselijke interventie voor nodig. Twee belangrijke vormen van bias zijn ambigüiteit en onzekerheid. Deze twee vormen zullen in onderstaande paragrafen 2.4-2.4.3 worden uiteengezet. Daarnaast voegen de ontwikkelaars ook ambigüiteit van presentatie toe (March, 2010). De app geeft namelijk een visuele verhalende vorm van de data, wat een andere manier van presenteren is dan hoe dit nu in het EPD gebeurt. Met de vorm wordt ook waarde aan de informatie gegeven, omdat het iets doet met de interpretatie en verwerking ervan. Dit zal toegelicht worden in 2.4.2.

2.4 Medische data: ambigüiteit en onzekerheid

“Despite significant advances in scientific knowledge and technology, ambiguity and uncertainty are still intrinsic aspects of contemporary medicine” (Luther & Crandall, 2011, p.799). Ambigüiteit en onzekerheid zijn onlosmakelijk verbonden met de medische praktijk en medische data (p.800). Voor de besluitvorming van een medicus alsook voor een decision-support systeem is het bepalen van de risico’s

en voordelen voor een individuele patiënt mogelijk wanneer alle variabelen en relaties beschikbaar zijn (Kappen, 2015, p.108). Er is echter in bijna elk geval sprake van ontbrekende informatie; de variabelen en relaties zijn niet precies duidelijk (p.108). Daar komt bij dat er vaak veel intolerantie voor ambiguïteit en onzekerheid is: patiënten willen graag zekerheid van diagnose en prognose (Luther & Crandall, 2011, p.799). Daarnaast kan het de uitvoering van het werk van de medicus beïnvloeden: zij die gevoelig zijn voor onzekerheid zullen meer tests of behandelingen uitvoeren om meer te kunnen uitsluiten. Dit kan nadelig zijn voor de kosten van het ziekenhuis, maar ook voor de patiënt die wellicht risico's ondervindt van bepaalde behandelingen (p.799). Daar komt nog eens bij dat mensen slecht zijn in het inschatten van kansen en kansverdelingen (Kappen et al., 2016), waardoor onzekerheid ook problematisch kan zijn. Voor dit onderzoek is het interessant hoe deze twee aspecten, ambiguïteit en onzekerheid, tot uiting komen in de decision-support applicatie UrSmartStatus en hoe ontwikkelaars en medici hier mee omgaan. Om de twee termen in relatie tot de medische praktijk wat meer te verkennen, is de volgende theoretische deelvraag opgesteld: *wat betekent de ambiguïteit en onzekerheid van medische data?* Eerst zal ambiguïteit worden besproken en daarna onzekerheid van medische data.

2.4.1 Ambiguïteit van medische data

Ambiguïteit kan omschreven worden als onnauwkeurigheid of dubbelzinnigheid (Luther & Crandall, 2011, p.799). Voor dit onderzoek wordt het omschreven als dubbelzinnigheid, omdat het gaat om de dubbelzinnigheid van medische data. Dat betekent bijvoorbeeld dat een oplopende bloeddruk voor een bepaald ziektebeeld ernstig kunnen zijn, maar voor een gezond persoon die net heeft hardgelopen geen alarmerend signaal is. De interpretatie van de gegeven data (oplopende bloeddruk) is dus contextafhankelijk en patiëntgebonden. Noordegraaf en Abma (2003) beschrijven in hun artikel *Management by measurement* de term ambiguïteit. Hoewel van toepassing voor managementpraktijken in de publieke sector, is hun beschrijving ook kenmerkend voor medische data. Zij stellen namelijk dat wanneer eenzelfde informatie geïnterpreteerd wordt, dit voor verschillende mensen een andere betekenis kan krijgen (p.861). Ambiguïteit wordt gedefinieerd als “an ongoing stream that supports several different interpretations at the same time” (Weick, 1995; March & Olsen, 1989 in Noordegraaf & Abma, 2003, p.861). Noordegraaf en Abma baseren zich op March en Olsen (1989) wanneer zij vier verschillende soorten ambiguïteit beschrijven. Medische data kan gezien worden als ambiguïteit van participatie: “who is present determines how things are interpreted” (p.861). Voor dit onderzoek zou de ‘who’ zowel de ontwikkelaar als de arts zijn: beide voegen interpretatie toe aan de beslissing. Deze ruimte voor interpretatie wordt ook wel *interpretive spaces* genoemd: bij signalen of situaties (in dit geval nieuwe medische data) weet je nooit precies wat er aan de hand is en wat er gedaan moet worden (p.862). Deze interpretatieruimte ligt bij de arts. Natuurlijk is elke situatie verschillend en zijn er patiënten waarbij het snel duidelijk is en data weinig ambigu is, maar ook patiënten waarbij de data wel omstreden is. Noordegraaf en Abma geven hier respectievelijk de termen ‘canonical practices’ en ‘non-canonical practices’ voor (p.863). In het eerste geval zijn situaties, medische data, bekend en de standaarden ervoor niet omstreden; er is weinig ambiguïteit. Er is weinig ruimte voor interpretatie en het is veelal duidelijk wat er gedaan moet worden. Voor algoritmes zou er min of meer een simpele ‘als X dan Y’ redenatie toegepast kunnen worden (pp.863-865). In het laatste geval zijn situaties onbekend en standaarden wel omstreden; er is veel ambiguïteit. Er moet onderzocht worden wat er aan de hand is, welke betekenis dat heeft en wat er gedaan moet worden (pp.863-867).

Schön (1987) stelt in zijn boek *Educating the Reflective Practitioner* dat ambiguïteit onderdeel is van de medische praktijk (Schön, 1987 in Luther & Crandall, 2011, p.800). De relaties tussen factoren en ziektes zijn soms duidelijk, maar er zijn er ook die onduidelijk blijven (Senger & O’Leary in Cohen et al., 2018, p.288). Hoe sommige factoren in relatie staan tot elkaar, vaak ook met complexe multivariabele relaties, is niet altijd duidelijk mede door gebrek aan onderzoek hiernaar (p.288).

Daarnaast zorgt ontbrekende data in de praktijk er ook voor dat sommige relaties niet kunnen worden uitgesloten. De vraag is hoe een algoritme binnen een decision-support systeem hier mee om kan gaan, als niet alle mogelijkheden eindig zijn (p.288).

2.4.2 Ambigüiteit van presentatie

Bij de waarde-creatie gaat het niet alleen om de ambigüiteit (en onzekerheid) van medische data, maar gaat het ook om de presentatie. Met de applicatie wordt de data namelijk ook op een heel andere manier gepresenteerd aan de artsen dan dat dit in het EPD wordt gedaan. Deze vorm van presentatie kan gezien worden als ambigüiteit van presentatie; een soort abstractere manier van presenteren. Dit idee komt van James March (1987; 2010). Hij stelt dat het ambigu presenteren van informatie ervoor zorgt dat er een rijkere betekenis aan kan worden gegeven om het begrip te vergroten (1987, p.165). March maakt daarbij een vergelijking met poëzie: “[...] *the essence of poetry lay in providing stimuli to the elaboration of meaning, rather than in providing unequivocal texts*” (p.165). Het begrip of de duiding kan namelijk groter zijn wanneer er geen eenduidige, volledig ingevulde dataset of beslissingsboom wordt gegeven, omdat er meer ruimte is voor eigen interpretatie (p.165). De data in de applicatie wordt namelijk weergegeven als een soort visueel verhaal en samenvatting van de patiënt. De kleuren geven geen harde, eenduidige betekenis aan de data, ze geven een gradatie aan van goed-slecht. Daarin ligt nog veel interpretatie voor de arts om te bepalen wat dat goed en slecht dan betekent. Daarvoor moet hij of zij de waardes bekijken die erachter liggen. Bovendien zijn verhalen en modellen voor mensen vaak simpeler te begrijpen dan complexe processen (March, 2010, p.44). Mensen zijn ontvankelijker voor simpele causaliteiten en heuristische die minder en simpelere informatie en berekeningen bevatten dan complexe analyses (p.45). Met een simpel en visueel overzicht van de gezondheid van de patiënt als eerste indruk, wordt het voor de arts gemakkelijker om een beeld te krijgen van de status van de patiënt. In plaats van dat hij of zij eerst allerlei data moet verzamelen om tot een algemeen beeld te komen (en daarbij vaak ook data verzamelt die kan bevestigen waar de arts al aan denkt), krijgt hij of zij eerst een globaal beeld van de patiënt en kan dan de gedetailleerde informatie bekijken. Mensen hebben namelijk een beperkte capaciteit voor analyse, waardoor ze vaak minder kritisch zijn op bewijs dat hun verwachting of overtuiging onderschrijft (p.45).

Met het ambigu presenteren van de informatie kan als het ware een communicatieprobleem opgelost worden tussen de mens en de data. Daarnaast wordt ook de gestructureerde data (feitelijke getallen) in de app gecombineerd met de ongestructureerde data (de informatie in verhaalvorm). In het EPD staan deze twee gescheiden, in de app worden ze als het ware geïntegreerd doordat de kleuren van de organen een waarde of kwaliteit aangeven van de data en daarmee vertellen hoe het met de patiënt gaat. In het EPD moet dit beeld zelf gecreëerd worden door de feitelijke data te combineren met wat er over de patiënt is geschreven door de artsen.

Er is dus sprake van twee soorten ambigüiteit, namelijk de ambigüiteit als complexiteit van de casus en ambigüiteit als abstractheid van presentatie van data zoals in de app gebeurt. Dit zit in de visuele verhalende vorm waarmee de informatie gepresenteerd wordt. Dit is een abstractere manier van presenteren dan het geven van feitelijke data, waardoor de manier van presenteren als meer ambigu kan worden gezien in de app dan hoe het in het EPD gepresenteerd wordt. Hierbij wordt verwacht dat dit een voordeel oplevert voor de interpretatie van artsen, omdat het beter aansluit bij hoe mensen denken. Bovendien geeft het veel ruimte voor de interpretatie van de arts, die nodig is gezien hun klinische logica die een algoritme niet heeft.

2.4.3 Onzekerheid van medische data

Onzekerheid kan worden omschreven als een conditie of feit met gebrek aan voorspelbaarheid (Luther & Crandall, 2011, p.799). Onzekerheid kan ook uitgedrukt worden in waarschijnlijkheid (Schwartz & Griffin, 1986, p.10). Het gaat dan om de waarschijnlijkheid dat een bepaalde diagnose klopt of de waarschijnlijkheid van een bepaalde prognose. De (statistische) waarschijnlijkheidsdistributie is dan wel duidelijk, maar de precisie van de waarde niet (Kappen, 2015, p.108). Als voorbeeld hiervan neemt Kappen de imperfectie van een diagnostische test. We weten de waarschijnlijkheid van de diagnose door de voorspellende waarde van de test, maar we weten niet of de test juist is (p.108). Als er sprake van onzekerheid is, betekent dit dus dat er niet met volledige zekerheid voorspeld kan worden hoe de gezondheid van de patiënt zal verlopen, omdat andere diagnoses en prognoses niet uitgesloten kunnen worden. Dit betekent bijvoorbeeld dat patiënt X 80% kans heeft dat hij ziekte A heeft, maar ook 20% kans heeft dat hij ziekte B heeft. Ziekte B kan niet uitgesloten worden, omdat er bijvoorbeeld een gebrek is aan een bepaald symptoom dat kenmerkend is voor ziekte A, waardoor het met zekerheid ziekte A zou zijn. Er is dus gebrek aan voorspelbaarheid van het verloop van de gezondheid van de patiënt, waarbij de waarschijnlijkheidsdistributie wel bekend is.

Medische besluitvorming gaat altijd gepaard met deze onzekerheid. Bij een beslissing voor een behandeling is er altijd onzekerheid over wat het resultaat zal zijn. De meest voor de hand liggende keus is niet altijd de beste. De onzekerheid wordt dan uitgedrukt in waarschijnlijkheid ('probability'). Dit gebeurt op een schaal van 0 tot 1 (of 0% tot 100%), waarin 1 een zekerheid van waarschijnlijkheid is en 0 een onmogelijke waarschijnlijkheid. Voor een medicus is de 1 (of 100%) niet haalbaar, er is nooit garantie of een behandeling of methode de juiste is (Schwartz & Griffin, 1986, p.10). De waarschijnlijkheid uitgedrukt op een schaal van 0 tot 1 is een statistische manier om een kans uit te drukken. Je kan uitrekenen hoe groot een kans is op iets. Maar er is ook subjectieve waarschijnlijkheid, wat in de medische wereld veel vaker wordt gebruikt. Dit is een subjectieve overtuiging in de waarschijnlijkheid dat iets zal gebeuren. De overtuiging kan wel worden afgeleid uit de informatie over de statistische waarschijnlijkheid (pp.10-11). Met het gebruik van een decision-support systeem die een (eventuele statistische) prognose geeft van de gezondheid van de patiënt kan ook de manier waarop er met onzekerheid (of waarschijnlijkheid) wordt omgegaan veranderen. Het kan er bijvoorbeeld voor zorgen dat artsen minder geneigd zijn de uitkomsten van recente patiënten mee te wegen of emoties mee te laten spelen, omdat de decision-support applicatie een 'objectievere' suggestie doet en deze zaken niet meeweegt.

Op basis van empirische onderzoeken wordt gesteld dat mensen niet houden van onzekerheid en hier vaak negatief tegenover staan (Berger, Bleichrodt & Eeckhoudt, 2013, p.559; Camerer & Weber, 1992, p.325; Carney et al., 2007, p.234; March, 2010, p.36). Dit kan invloed hebben op de besluitvorming van medici, bijvoorbeeld bij de keuze voor behandelingen of extra onderzoek (p.559; Carney et al., 2007, p.234). Uit het onderzoek van Berger, Bleichrodt en Eeckhoudt blijkt dat wanneer er onzekerheid bestaat bij de diagnose van een patiënt, de negatieve houding ten opzichte van onzekerheid ervoor zorgt dat de medicus eerder kiest voor behandeling. Het omgekeerde effect gebeurt wanneer er onzekerheid is over de effecten van een behandeling. Dan leidt een negatieve houding ten opzichte van onzekerheid tot een mindere neiging om de behandeling te kiezen (p.559). In het onderzoek van Carney et al. (2007) zorgde ongemak van radiologen bij onzekerheid van mammogrammen¹ ervoor dat zij vaker meer onderzoek deden dan radiologen die minder ongemak ervaren bij onzekerheid. De houding ten opzichte van onzekerheid heeft dus invloed op de besluitvorming van medici. Daarnaast speelt het inschatten van

¹ Een mammogram is een röntgenfoto van de borstklier voor het opsporen van borstkanker. Via: <https://www.ziekenhuis.nl/onderzoek-diagnose/m/mammografie/item314>

kansen ook een rol. Kahneman en Tversky (1972) stellen dat mensen in het beoordelen van kansen in onzekere situaties niet de principes van waarschijnlijkheid hanteren (p.430-431). Een oorzaak hiervan is dat de waarschijnlijkheidsdistributie vaak niet overeenkomt met de intuïtie over de situatie, waardoor er een verschil is in objectieve en subjectieve waarschijnlijkheid (p.431). De objectieve kans wordt dan vervangen door heuristieken, bijvoorbeeld door representativiteit. Gebeurtenissen worden als waarschijnlijker geschat voor groep A wanneer deze populatie representatiever is dan groep B ten opzichte van de groep waarmee het wordt vergeleken (p.431).

2.5 Besluitvorming in de medische praktijk

In deze paragraaf zal de deelvraag *hoe werkt rationele en medische besluitvorming?* worden beantwoord. Besluitvorming is een essentieel onderdeel van de medische praktijk. Er zal eerst een korte introductie worden gegeven op waar de medische besluitvorming vandaan komt, namelijk het medisch oordelen en redeneren. Vervolgens wordt uiteengezet hoe rationele besluitvorming werkt, waarin de basis ligt voor algoritmische besluitvorming en decision-support systemen. Vervolgens wordt de koppeling gemaakt met medische besluitvorming. Een belangrijke factor in deze besluitvorming is de medische voorspelling met de verwachte utiliteit. Hier zal ten slotte bij worden stilgestaan.

2.5.1 Medisch oordelen en redeneren

Het werk van een medicus bevat constant het vormen van oordelen en het maken van beslissingen. De basis hiervoor werd gelegd in het begin van de 19^e eeuw (Schwartz & Griffin, 1986, p.5). Artsen behandelden iedereen die medische klachten had, maar zagen dat onbehandelde patiënten soms uit zichzelf beter werden. Dat was de aanleiding om diagnoses te gaan stellen en voorspellingen te maken over welke patiënten beter zouden worden en welke niet (p.5). Het observeren van en luisteren naar patiënten en hun symptoombeschrijvingen, het koppelen van deze signalen aan mogelijke ziektes en vervolgens verder observeren en testen wordt door Schwartz en Griffin ook wel de ‘essentie van wetenschappelijk redeneren’ genoemd (p.5). De empirische, inferentiële methode geeft de geneeskunde een wetenschappelijk karakter. Maar, zo stellen Schwartz en Griffin, de geneeskunde blijft ook een kunst (‘art’) (p.6). De kunst van geneeskunde is namelijk het bedachtzaam toepassen van de wetenschap. De auteurs maken dit duidelijk aan de hand van een voorbeeld waarin een huisarts zes kinderen in zijn praktijk krijgt die buikpijn hebben na een picknick. De voor de hand liggende diagnose leek voedselvergiftiging. Echter bleek na zorgvuldige inspectie van de huisarts dat vijf kinderen een voedselvergiftiging hadden en één kind een blindedarmonsteking (pp.6-7). Zo een arts beheerst de kunst van de geneeskunde en past de wetenschap bedachtzaam en vaardig toe. Dit medisch oordelen en redeneren, de ‘kunst’ van geneeskunde, is steeds meer onderwerp van aandacht geworden in de wetenschap (p.7) en is van groot belang bij de oordeelsvorming en besluitvorming van medici.

De komst van decision-support systemen kan gezien worden als onderdeel van een ‘rationaliseringslag’ door *evidence-based medicine*; de visie dat het handelen en de beslissingen van medici gebaseerd zijn op wetenschappelijk bewijs. Het handelen wordt gerationaliseerd op basis van ‘het beste wat we weten’ waardoor het ‘bewezen effectief’ wordt (Van Bodengraven, 2018). De term ‘rationeel’ betekent in deze context een op feiten gebaseerd logisch stappenplan, waarbij enkel data wordt gebruikt die afkomstig is uit medische gegevens (lab-uitslagen, bloeddrukmetingen, etc.) en niet uit de ervaring en intuïtie van de arts. Het is de vraag hoe de ‘kunst van geneeskunde’ zich hiertoe verhoudt wanneer decision-support systemen zoals UrSmartStatus hun intrede doen.

2.5.2 Besluitvorming

Om de werking van zowel rationele als medische besluitvorming te begrijpen is het noodzakelijk eerst wat meer te weten over besluitvorming op zichzelf. Besluitvorming is “a process that chooses a preferred

option or a course of actions from among a set of alternatives on the basis of given criteria or strategies” (Wang & Ruhe, 2007, p.73). Besluitvorming is een mentaal basisproces dat elke paar seconde, bewust of onbewust, plaatsvindt (p.74). Er zijn fundamentele strategieën voor besluitvorming: intuïtie, empirie, heuristiek en rationaliteit (p.76).

Daarnaast zijn er twee verschillende manieren, ook wel paradigma's genoemd, waarop er naar besluitvorming kan worden gekeken: descriptief en normatief (Peterson, 2009, p.3; Wang & Ruhe, 2007, p.74). Descriptieve theorieën zijn gebaseerd op empirische onderzoeken naar gedrag, terwijl normatieve theorieën ervan uitgaan dat er een rationele besluitvormer bestaat die volgens bepaalde regels of axiomen besluiten maakt. De twee benaderingen hebben het startpunt van besluitvorming wel gemeen. Ze gaan er namelijk allebei vanuit dat een besluitvormingsproces start met de overtuigingen of verlangens van de besluter (Peterson, 2009, p.4). Rationele besluitvorming is het onderwerp van de (*normative*) *decision theory* (Peterson, 2009, p.1). Voorbeelden van theorieën die hieronder vallen, zijn de theorie van *expected utility* en de Bayesiaanse theorie (Wang & Ruhe, 2007, p.74). Hierbij wordt ervanuit gegaan dat er te onderscheiden uitkomsten zijn die een bepaalde waarde hebben of krijgen, en daaruit rationeel logisch beredeneerd kan worden dat een bepaalde uitkomst de beste waarde heeft en dus de beste beslissing is. De opties moeten waarde (een besluitvormingsdoel en selectiecriteria) krijgen, maar vervolgens kan de keuze gemaakt worden door zowel mens of machine: het is immers enkel een berekening van de beste waarde (p.74). De basis van decision-support systemen gaat dan ook uit van rationele besluitvorming: er wordt een rationele logica ingebouwd waarin waarde wordt gegeven aan bepaalde factoren.

2.5.3 Rationele besluitvorming

Rationele besluitvorming kent twee soorten strategieën en criteria: statisch en dynamisch (Wang & Ruhe, 2007, p.75). Statisch houdt in dat omgevingsverandering onafhankelijk van de besluitvormer is. Dynamische situaties zijn situaties waarin de omgeving interactief werkt op de besluiten die worden gemaakt (p.77). Rationele besluitvorming laat zich kenmerken door drie essenties: doelen, alternatieve keuzes en selectiecriteria of –strategieën. Dit wordt ook wel het axioma van keuze genoemd (Lipschutz, 1967 in Wang & Ruhe, 2007, p.74). Volgens dit axioma zijn de besluitvormers de uitvoerders van een besluitvormingsproces en wanneer de drie essenties gedefinieerd zijn, kan het besluitvormingsproces zowel door een mens als door een intelligent systeem worden gedaan. Dit axioma ligt dan ook vaak ten grondslag aan de implementatie van decision-support systemen (p.74). Met het idee dat mensen beslissingen moeten maken in complexe, onzekere en dynamische omgevingen, kunnen decision-support systemen nuttig zijn (p.83).

Een rationele beslissing hoeft niet altijd een juiste beslissing te zijn: “A decision can be rational without being right and right without being rational” (Peterson, 2009, p.4). Een beslissing is rationeel wanneer de besluitvormer het besluit kiest waar hij of zij de meeste reden toe had. Een beslissing is juist wanneer de uitkomst tenminste net zo goed is als de andere mogelijke uitkomsten van andere beslissingen (p.5).

Bij rationele besluitvorming zijn risico en onwetendheid (*ignorance*) belangrijk voor de afweging van mogelijke uitkomsten. Bij besluiten met risico's is de waarschijnlijkheid van de mogelijke uitkomsten bekend bij de besluitvormer en bij besluiten met onwetendheid is de waarschijnlijkheid van de mogelijke uitkomsten niet bekend of niet-bestaand. De term onzekerheid wordt daarbij gebruikt voor onwetendheid of voor zowel risico's en onwetendheid (pp.5-6). Bij besluiten met risico's wordt vaak het principe van het maximaliseren van de verwachte waarde (*maximising expected value*) gebruikt.

2.5.4 Medische besluitvorming

Het besluitvormingsproces van een medicus, in tegenstelling tot enkel rationele besluitvorming zoals de decision theory beschrijft, bestaat uit zowel intuïtieve en ervaringskennis (*tacit*) als rationele logica. Het handelen van een medicus is een ontwerpproces, zo stelt Donald Schön (1991). Hij duidt dit ook wel aan met de term ‘reflection-in-action’ (p.170) waar Van Os (2009) een heldere beschrijving van geeft:

“Pas als het probleem juist gedefinieerd wordt, is ook een zinvolle oplossing te vinden. Het probleem juist definiëren vraagt reflectie op de concrete situatie waarin het probleem moet worden opgelost. De professional reflecteert op de specifieke situatie, formuleert voor zichzelf het specifieke probleem en de bijbehorende probleemoplossende actie. Reflection-in-action vraagt een grote mate van vaardigheid om tegelijkertijd te reflecteren en handelen” (Van Os, 2009, p.101).

De medicus gebruikt voor dit reflecteren en handelen uiteraard wel zijn wetenschappelijke, ‘technical’ basis (Schön, 1991, p.170; Wilensky, 1964). Hij maakt voor zijn besluitvorming dus gebruik van zijn autonomie. Dit betekent dat de medicus de capaciteit heeft tot het maken van onafhankelijke, zelfstandige keuzes en zich daarvan bewust is. Hij kan op deze keuzes reflecteren en handelen op basis van een beredeneerde beslissing die voortkomt uit de inhoudelijke vakkennis maar ook uit zijn kritisch vermogen tot reflectie (Van Oorschot et al., 1995, pp.16-17). Professionele autonomie kan daarmee gezien worden als een essentiële kernkwaliteit van de medicus (p.17). De medicus gebruikt dus zijn professionele autonome ruimte om tot een oordeel te komen over de juiste manier van handelen in het belang van de individuele patiënt.

Een belangrijk onderdeel van de medische besluitvorming is het doen van voorspellingen. In Schwartz & Griffin (1986) wordt er onderscheid gemaakt tussen klinische en statistische voorspelling (p.7). Hierbij wordt dezelfde data op een andere manier gecombineerd en geëvalueerd (p.8). Bij statistische voorspelling worden statistische regels of formules toegepast om de statistische waarschijnlijkheid op een uitkomst te voorspellen (pp.7-8). Dit zou als een rationele logica beschouwd kunnen worden. Bij klinische voorspelling wordt de intuïtie en ervaring (*tacit*) gebruikt met als basis de geschiedenis van de patiënt, huidige symptomen en uitslagen van het lab om de waarschijnlijkheid te voorspellen (pp.7-8). Voor beide zijn kritiekpunten te leveren: klinische voorspelling kampt met onzekerheid en bij statistische voorspelling is de relatie tussen diagnose en behandeling nooit volmaakt (p.8).

Naast (statistische) waarschijnlijkheid speelt utiliteit ook een grote rol in de medische besluitvorming. Bij de besluitvorming wordt dan ook de ‘voorkeur’ van de patiënt meegenomen. Het gaat dan om bijvoorbeeld kosten, pijn en kwaliteit van leven. Een behandeling kan namelijk voor een individueel geval heel waardevol zijn, de kwaliteit van leven kan flink vooruitgaan (Schwartz & Griffin, 1986, p.11). Maar bij een andere patiënt kan die vooruitgang een stuk minder zijn. Dit kan meegenomen worden in de besluitvorming.

Met de komst van een decision-support systeem kan de besluitvorming van de arts veranderen doordat wellicht rationele besluitvorming een grotere rol gaat spelen. De vraag is of en hoe dit invloed heeft op de medische besluitvorming waarin intuïtie, ervaring en utiliteit ook een rol spelen.

2.6 Conclusie en verwachtingen

Uit het theoretisch kader is gebleken dat decision-support systemen medici kunnen ondersteunen in hun besluitvorming. Veel van de systemen kennen echter nog veel verbeterpunten waardoor de implementatie ervan discutabel kan zijn. Daarnaast kennen de algoritmes die de basis vormen van deze decision-support systemen ethische bezwaren. Zo wordt vaak gedacht dat algoritmes objectieve logica’s gebruiken, maar deze zijn wel waarde-geladen omdat ze door ontwikkelaars zijn gemaakt. Algoritmes

kunnen zo complex zijn dat ze onnavolgbaar zijn voor de gebruikers en daarmee een ‘black-box’ vormen. De legitimiteit van de output is daardoor soms niet te achterhalen en de interpretatie ervan lastig. Met de context van medische besluitvorming kan dit voor problemen zorgen, gezien de gevolgen die medische besluiten kunnen hebben op de gezondheid en levens van patiënten. Ook speelt de ambiguïteit en onzekerheid van medische data een belangrijke rol in het ontwikkelen en interpreteren van een algoritme. Dit zorgt er namelijk voor dat de input voor het algoritme niet eenduidig is, dus de vraag is hoe dit dan verwerkt en gecommuniceerd wordt naar de gebruiker. De gebruiker (de medicus) en het algoritme hebben namelijk een verschillende logica in besluitvorming: een algoritme maakt gebruik van rationele besluitvorming, terwijl medici ook een tacit logica gebruiken: subjectieve, klinische overtuigingen.

In de app UrSmartStatus wordt de besluitvorming gerationaliseerd, maar is er wel sprake van ambiguïteit en onzekerheid van medische data. Het is de vraag hoe ontwikkelaars hier mee omgaan. Daarnaast gebruiken medici zowel rationele als klinische logica. De vraag is hoe hun besluitvormingsproces verandert met het gebruik van de app UrSmartStatus. Dit theoretisch kader geeft de basis voor het empirisch onderzoek: hoe krijgt de medische data waarde in de decision-support app UrSmartStatus in het UMC Utrecht?

Op basis van het theoretisch kader kunnen er ook verwachtingen worden geformuleerd ten aanzien van het empirisch onderzoek. Ten eerste wordt verwacht dat het ontwikkelen van de decision-support app UrSmartStatus in het UMC Utrecht niet gemakkelijk zal zijn en (ethische) consequenties kan hebben. De app is nog in ontwikkeling en nog niet geïmplementeerd, dus dit onderzoek kan ook bijdragen aan het overzien van de (on)mogelijkheden en inzicht geven in de processen van waarde-creatie. Met de uitkomsten van de onderzoeken naar decision-support systemen kan rekening worden gehouden tijdens de dataverzameling. Zo zou de geschiktheid voor individuele patiënten en specifieke klinische omstandigheden te maken kunnen hebben met de bias die medische data kent. Omdat een algoritme een rationele logica kent, kan bias lastig te verwerken zijn, als het al mogelijk is. Daarnaast zou voor het verzamelen van de data bij medici de welwillendheid van het gebruik van CDSS een rol kunnen spelen, dus follow-up vragen hiernaar zijn belangrijk. De verwachting voor het experiment met de artsen is dat het decision-support systeem nog niet naadloos zal aansluiten bij het werkproces van de artsen doordat zij niet gewend zijn aan het werken met een dergelijk nieuw systeem, wat voortkomt uit de evaluatiestudies van en onderzoeken naar klinische decision-supportsystemen en de discrepantie tussen de rationele logica van de applicatie en de klinische logica van artsen.

Ten tweede wordt verwacht dat het voor de ontwikkelaars een complexe taak is om goed om te gaan met de context van medische besluitvorming ten aanzien van de algoritmes. De app heeft als doel de medicus te ondersteunen, maar kan ook snel de medicus (onbedoeld) in een bepaalde richting sturen. Wanneer data niet eenduidig is of kan zijn, kan het implicaties hebben voor de medische praktijk. Voor de ontwikkelaars is er dus de taak om genoeg ruimte voor de medicus over te laten voor zijn of haar professionele, klinische besluitvorming. Het algoritme zou idealiter de imperfecties van de menselijke besluitvorming weg moeten nemen die een algoritme beter kan (zoals kansen inschatten), maar niet interveniëren met de kritische blik die een medicus kan geven door zijn of haar ervaring en intuïtie. Verwacht wordt dat de applicatie de medicus kan helpen in de rationele logica door beter (en sneller) aan te sluiten bij hoe mensen denken, maar dat de klinische logica bij de arts blijft. Hoe ontwikkelaars omgaan met deze twee logica's en hun logica communiceren naar de artsen, zal blijken uit het empirisch onderzoek. Ook zal blijken hoe medici interpretaties maken over de data en in hoeverre de waarde-creatie van de ontwikkelaars een rol daarin speelt.

Ten slotte zijn er ook verwachtingen ten aanzien van de ambiguïteit en onzekerheid van medische data voor de resultaten van het onderzoek. Met betrekking tot de ambiguïteit wordt verwacht dat het voor ontwikkelaars erg lastig is hoe ze hier mee om moeten gaan. Hoe kan een algoritme beredeneren of de gegeven data incompleet is om een gedegen diagnose te stellen en de ernst hiervan weer te geven? En moet het algoritme hier wel iets mee doen, of blijft dit de taak van de arts? De verwachting ten aanzien van de ambiguïteit van de presentatie is dat dit zal helpen in het interpretatieproces van de arts. Doordat de applicatie de data meer ambigu presenteert dan dat nu wordt gedaan, zal dit beter aansluiten bij hoe artsen denken. Hierdoor zullen ze met minder moeite een interpretatie maken van de gezondheid van de patiënt en zal dit wellicht ook sneller gaan. Door de overzichtelijkheid van de presentatie in de app wordt verwacht dat dit de arts helpt in een zo compleet mogelijk beeld, waardoor de diagnose uitgebreider kan zijn dan de diagnose die gesteld wordt bij het gebruik van de huidige systemen. Daarnaast is de verwachting dat de onzekerheid van data juist goed kan worden weergegeven in de app waardoor medici dit beter kunnen interpreteren en hier minder subjectiviteit in zal voorkomen. Een computer kan goed omgaan met statistische waarschijnlijkheid. Wanneer dit op de juiste manier wordt gecommuniceerd, kan dit de arts helpen in zijn of haar besluitvorming. De verwachtingen ten aanzien van de interpretatie van de artsen zijn voor het experiment ook geformuleerd als hypotheses, deze zijn te vinden in het methodehoofdstuk onder 3.5.5.

3. Methode

Voor het beantwoorden van de hoofdvraag *Hoe krijgt medische data waarde in de decision-support applicatie UrSmartStatus in het UMC Utrecht?* en de deelvragen zijn er twee componenten voor de dataverzameling en -analyse: de waarde-creatie van de ontwikkelaars en de interpretatie van de medici. Deze data verschilt van elkaar, omdat het twee groepen bevat die op een andere manier betekenis geven aan medische data. Bij de ontwikkelaars gaat het om inzicht in hoe zij betekenis geven aan onder andere ambiguïteit en onzekerheid van medische data en hoe dit vorm krijgt in de app. Bij de medici gaat het om inzicht in hun interpretatieproces; hoe gaan zij om met de betekenis die gegeven is aan de medische data in de app? En is de interpretatie afhankelijk van de mate van ambiguïteit in patiëntcasussen? In de deelvragen is dit onderscheid al gemaakt. Deze zijn gebruikt als uitgangspunt voor de keuze van de methode; er zijn dan ook verschillende methodes gebruikt voor de dataverzameling en data-analyse. Voor de dataverzameling bij de ontwikkelaars is gekozen voor interviews, voor de medici is gekozen voor een (kwalitatief) experiment.

3.1 Case study

In dit onderzoek gaat het specifiek over de ontwikkeling van de decision-support app UrSmartStatus in het UMC Utrecht. De resultaten hebben dus ook alleen betrekking op deze specifieke app, en geen andere decision-support systemen. Deze specifieke decision-support app is niet vergelijkbaar met andere decision-support systemen en tevens in ontwikkeling, wat de onderzoekssituatie bijzonder maakt. Om deze reden is er dan ook sprake van een case study (Bryman, 2012, pp.66-67). In deze situatie wordt er ook wel gesproken over een ‘extreme’ of ‘unieke’ case, wat gebruikelijk is in klinische onderzoeken (Yin, 2009 in Bryman, 2012, p.70). Dat betekent dat de resultaten van het onderzoek betrekking hebben op deze specifieke setting. Dat wil niet zeggen dat de resultaten niet een bredere relevantie kunnen hebben. Het gaat namelijk niet om het gebruik van de applicatie, maar om hoe de data waarde krijgt. Dit waarde-creatie proces zal ook bij andere decision-support applicaties plaatsvinden waardoor dit onderzoek kan bijdragen aan inzicht in dit proces. Daarnaast kan vooral ook het experimentele gedeelte van dit onderzoek gezien worden als een eerste opzet voor het onderzoeken van de implicaties van nieuwe technologie op het besluitvormingsproces van artsen.

Case beschrijving

De applicatie UrSmartStatus gaat een actuele visuele weergave geven van de gezondheid van patiënten, weergegeven in een overzicht van de organen waarbij in één oogopslag te zien moet zijn met welke organen het goed gaat en met welke niet. Dit is aangeduid door de kleuren groen, oranje en rood, waarbij groen aanduidt dat het min of meer goed gaat met het orgaan en oranje en rood gradaties van verslechtering zijn. De organen kleuren lichter wanneer er op een manier data ontbreekt en het orgaan wordt grijs wanneer er geen informatie over is. Op de organen kan geklikt worden waardoor er een scherm openklapt met een historisch tijdspad en de bijbehorende waarden. Daarnaast gaat de app een prognose geven van het verloop van de gezondheid op basis van probabiliteit (waarschijnlijkheid), maar dit zal in het experiment niet van toepassing zijn. De achterliggende gedachte hierbij is dat de interface gebruikersvriendelijk en intuïtief is in het weergeven van de patiëntdata. Het doel kan omschreven worden als ‘real-time data-driven storytelling’: de applicatie vertelt het verhaal van de patiënt in visuele vorm op basis van de beschikbare data (Kappen, 2019). Hierbij is de ambitie om medici sneller besluiten te kunnen laten maken en wellicht in een later stadium ook betere beslissingen. De app zal in eerste instantie complementair zijn aan het EPD waarin data geordend is op input. Dat betekent dat er bijvoorbeeld in het ene tabblad de lab-uitslagen staan weergegeven en in het andere tabblad de bloeddruk uitslagen. De data is eenduidig, maar is logisch gerangschikt voor de dataleveraars (o.a. laboranten), niet voor de datagebruikers (artsen) (Kappen, 2019). Hierbij wordt gesuggereerd dat het voor de gebruikers logischer zou zijn om op inhoudsniveau de informatie te krijgen, dus bijvoorbeeld ingedeeld

op organen. De precieze uitwerking en bepaling is nog in ontwikkeling. In de bijlages F en G zijn printscreens te zien van de applicatie zoals die gebruikt is in het experiment.

Deze case study is interessant voor dit onderzoek, omdat het inzicht geeft in het ontwikkelingsproces en kan tegelijkertijd dienen als *usability test*. Het hele proces van waarde-creatie bij het gebruiken van een decision-support systeem kan in kaart worden gebracht, omdat er zowel gekeken wordt naar de waarde die de ontwikkelaars er in stoppen, als de waarde die de gebruikers (de medici) eruit interpreteren. Dit kan informatie bieden over de effectiviteit, functionaliteit en wenselijkheid van de decision-support applicatie.

3.2 Verkennende fase

Om meer inzicht te krijgen in de klinische praktijk en wat het project rondom de ontwikkeling van de applicatie inhoudt, zijn er in een verkennende fase gesprekken geweest met de projectleiders en presentaties van de applicatie bijgewoond. De gesprekken gingen zowel over de applicatie als over de klinische praktijk en patiëntcasussen. Daarnaast is er een evaluatierapport gedeeld over de voortgang van het project en heeft er een meeloopochtend plaatsgevonden waarbij inzicht is gekregen in de dagelijkse praktijk van de IC op de kinderafdeling. Een verslag hiervan is te vinden in bijlage A.

3.3 Dataverzameling ontwikkelaars

Om inzicht te krijgen in de waarde-creatie van de ontwikkelaars, vooral met betrekking tot ambiguïteit en onzekerheid van medische data, is de volgende deelvraag opgesteld: *hoe wordt waarde toegekend aan medische data door de ontwikkelaars van UrSmartStatus?*

Daarin gaat het met name om hoe zij omgaan met ambiguïteit en onzekerheid van medische data, omdat daarin vooral waarde toegekend wordt aan data. Hoe geven zij betekenis aan de interpretatie van medische data door het algoritme? Hoe wordt dit gecommuniceerd in de app? Ter verheldering een voorbeeld: een patiënt heeft verschillende aandoeningen (co-morbiditeit) en komt het ziekenhuis binnen met meerdere klachten. Voor de juiste behandeling is het van belang te achterhalen uit welke aandoening deze klachten voortkomen. Kan een algoritme dit bepalen en hoe wordt dit vertaald naar de actuele status van de patiënt in de app? Daarnaast is het stellen van een diagnose ook van belang voor het kunnen maken van een prognose. Bepaalde symptomen of klachten kunnen voor de ene aandoening ernstige gevolgen hebben voor de gezondheid van de patiënt, maar bij een andere aandoening niet direct tot gevaarlijke situaties leiden. Hoe wordt dit vertaald naar de prognose van de gezondheid van de patiënt, hoe wordt omgegaan met de onzekerheid van die prognose?

Om inzicht te krijgen in dit proces van waarde-creatie aan de ontwikkelaarskant, zijn semi-gestructureerde interviews gehouden met de ontwikkelaars aan de hand van een topiclijst (Bryman, 2012, p.470). Hierbij ging de dataverzameling om het verkrijgen van inzicht in hoe de waarde-creatie tot stand komt en welke redenen er zijn voor bepaalde beslissingen. Deze interviews waren kwalitatief van aard, dus boden veel ruimte om de inhoud van het interview gedeeltelijk te baseren op wat de respondent vertelde (p.470). Het proces van ontwikkeling gebeurt niet individueel maar dit doen de ontwikkelaars samen. Daarom zijn de interviews gedeeltelijk met twee ontwikkelaars tegelijkertijd gehouden om inzicht te krijgen in hoe zij gezamenlijk tot bepaalde besluiten komen. Er zijn in totaal vijf interviews gehouden, waarvan twee interviews met twee ontwikkelaars. Eén van de interviews was met één van de oprichters en projectleider van de applicatie, die het overzicht en de visie bewaart voor het project. De interviews werden opgenomen als audiomateriaal waarvoor toestemming is gevraagd aan de betreffende respondenten.

3.4 Data-analyse ontwikkelaars

Het audiomateriaal van de interviews is getranscribeerd en vervolgens gecodeerd (Bryman, 2012, p.568). Bij het coderen is eerst open, vervolgens axiaal en ten slotte selectief gecodeerd (p.569). Bij al deze fases werd specifiek gelet op waarde-creatie en betekenisgeving van en omgang met medische data, vooral met betrekking tot ambiguïteit en onzekerheid van medische data. De topiclijst in bijlage B diende hiervoor als leidraad. Bij het open coderen werden er concepten uit de transcripten gehaald, die vervolgens tot categorieën zijn gelabeld (p.569). Het axiaal coderen ging om het vinden van overkoepelende thema's, uitingen of concepten. Bij het selectief coderen werd de argumentatie opgebouwd door relaties te leggen tussen de codes (p.569). Voor het coderen is gebruikgemaakt van NVivo.

3.5 Data verzameling medici: experiment

Om inzicht te krijgen in de waarde-creatie van medici, of ook wel hun interpretatie van de gegevens uit de app en hun bijbehorende besluitvormingsproces, is de volgende deelvraag opgesteld: *hoe wordt waarde toegekend aan informatie uit de applicatie UrSmartStatus door medici?*

Hiertoe is er met zes artsen binnen de kindergeneeskunde van het Wilhelmina Kinderziekenhuis een experiment uitgevoerd. Hierbij zijn twee patiëntcasussen ontwikkeld waarin een fictieve maar realistische situatie beschreven werd van patiënten die al een aantal dagen in het ziekenhuis liggen waarbij zich een verandering in hun gezondheidstoestand voordoet. Deze casussen verschilden in mate van complexiteit betreffende de ambiguïteit van de casus. De casussen zijn zowel verwerkt in een samengevoegde versie van het EPD waar de artsen nu in werken (HiX en Metavision) als in een versie van de applicatie UrSmartStatus. Aan de hand van de *think-aloud* methode is aan de artsen gevraagd hardop na te denken over wat zij in de beschreven situaties zouden doen; tot welke differentiaal diagnose zij zouden komen en welk beleid ze zouden voeren. De *think-aloud* methode is een vorm van *process-tracing* en houdt in dat het gedachteproces van de artsen in het besluitvormingsproces wordt gevolgd (Lundgrèn-Laine & Salanterä, 2010; Todd & Benbasat, 1987). Na de *think-aloud* sessie werden een aantal korte follow-up vragen gesteld, deels gebaseerd op de bevindingen tijdens de *think-aloud* sessie. Deze zijn te vinden in bijlage E. Bij de *think-aloud* techniek gaat het om rijke data van een kleine groep respondenten (Fonteyn, Kuipers, & Grobe, 1993, p.432). Het onderzoek gaat niet om het generaliseren over een bepaalde populatie en de informatie uit één individu is al heel rijk, dus een kleine groep respondenten is voldoende om iets over de verkregen data te kunnen zeggen (p.432).

3.5.1 Patiëntcasussen in EPD en app

De casussen zijn ontwikkeld in samenwerking met twee artsen van het UMC Utrecht die ook verbonden zijn aan het UrSmartStatus project. In de casussen is er gevarieerd in de mate van ambiguïteit om te onderzoeken hoe medici dit interpreteren. Er is voor gekozen om op ambiguïteit te variëren en niet (ook) op onzekerheid, omdat onzekerheid zou gaan over de prognose die in de app wordt gegeven. Deze prognose is voor het experiment weggelaten, om de vergelijking tussen het gebruik van de app en het gebruik van het EPD gelijkwaardig te maken. Een dergelijke prognose zit namelijk niet in het EPD en er zou geen goed vergelijk gemaakt kunnen worden wanneer dit in het geval van gebruik van de app wel gebruikt zou kunnen worden om tot een diagnose te komen. De patiëntcasussen zijn te vinden in bijlages F en G. Om de interpretatie van de artsen te kunnen duiden werden de casussen zowel aangeboden in het EPD als in de applicatie om een vergelijk te maken. Door te kijken naar het verschil van interpretatie en besluitvormingsproces tussen het gebruik van het EPD en het gebruik van de app kon er een uitspraak worden gedaan over hoe de waarde-creatie van de ontwikkelaars geïnterpreteerd werd en wat het gebruik van de app doet met de interpretatie van de medici. Aangezien het om een

fictieve casus gaat en de app nog niet zo functioneert dat die gebruikt kan worden en een fictieve casus in het EPD aanmaken geen optie was, werden zowel het EPD als de app in pdf aangeboden. Dit betekende dat zowel het EPD als de applicatie een nagemaakte versie waren van de echte, maar in uiterlijke weergave niet afdeden aan de echte versies. Voor de haalbaarheid van het experiment zijn de relevante tabbladen voor de casus in het EPD aangemaakt en een samengevoegde versie van HiX en Metavision gemaakt. Voor de applicatie is gebruikgemaakt van het tonen van de actuele status van de patiënt, met de optie om ook HiX en Metavision te kunnen bekijken. De mogelijkheid om een prognose te zien is in de applicatie weggelaten om de twee systemen gelijkwaardig te maken. Door middel van hyperlinks kon er doorgeklikt worden op dezelfde manier als dit in de echte versies van het EPD en de applicatie zou kunnen.

3.5.2 Think-aloud methode

Voor de dataverzameling van de interpretaties van de artsen is gebruikgemaakt van de *think-aloud* techniek gebaseerd op het onderzoek van Lundgrèn-Laine en Salanterä (2010) als vorm van *process-tracing* (Todd & Benbasat, 1987). Het idee van de think-aloud techniek is dat participanten wordt gevraagd om wat ze denken te verbaliseren, dus hun gedachteproces hardop uit te spreken tijdens het oplossen van een probleem of het uitvoeren van een taak zoals het nemen van een besluit (Fonteyn, Kuipers, & Grobe, 1993, p.430; Lundgrèn-Laine & Salanterä, 2010, p.568). De onderzoeker observeert de participant en neemt het als audio of video op (p.568). De think-aloud techniek van Ludgrèn-Laine en Salanterä (2010) komt overeen met wat Todd en Benbasat (1987) noemen een ‘verbal protocol analysis’ (p.496). Een protocol is “a verbal process, expressed aloud, which reveals the “step-by-step” progression of a person’s problem-solving ability” (Ludgrèn-Laine & Salanterä, 2010, p.566). De think-aloud techniek kan inzicht geven in het gedachteproces van de participant en daarmee informatie verschaffen over het redeneringsproces, welke informatie gebruikt wordt en hoe die gestructureerd is in het werkgeheugen om een probleem op te lossen of een besluit te nemen (Fonteyn, Kuipers, & Grobe, 1993, p.430). Om verwarring te voorkomen wordt de term think-aloud techniek gebruikt voor de manier van data verzameling: het hardop denken van de medici. Protocol analyse als term wordt gebruikt voor de data-analyse.

Het hardop laten uitspreken van de stappen die de medici zetten is gedaan aan de hand van zogenoemde ‘concurrent’ protocollen, en niet ‘retrospective’ protocollen (hierna: techniek) (Todd & Benbasat, 1987, p.497). Hiermee wordt bedoeld dat bij de concurrent techniek het gedachteproces tijdens het besluitvormingsproces werd uitgesproken. Dit is gedaan door neutraal te bevragen, en niet gestructureerd (p.498). Bij neutraal bevragen wordt er aan de participant instructie gegeven om het proces te beschrijven, zonder dat de onderzoeker specifieke vragen stelt (p.498). Er werd gevraagd om constant, bij iedere stap en iedere gedachte hardop na te denken. Bij elke stilte van ongeveer 10 seconden werd gevraagd om hardop te blijven nadenken. Er is verder geen interventie geweest tussen de participant en onderzoeker om het gedachteproces zo min mogelijk te storen, enkel het noodzakelijk beantwoorden van vragen van de participant zoals ‘Kan ik dan hier op klikken?’

Tijdens de sessie is door de onderzoeker aantekeningen gemaakt voor korte follow-up vragen na het hardop denken voor eventuele verheldering van het gedachteproces. Daarnaast zijn er ook van tevoren vragen opgesteld om bijvoorbeeld de attitude en het vertrouwen van de participant ten opzichte van de applicatie in beeld te brengen. Deze vragenlijst is te vinden in bijlage E. De gehele sessie is opgenomen als audiomateriaal met schermopname waarvoor toestemming is gevraagd.

3.5.3 Experiment operationalisatie

<i>Arts</i>	<i>Eerste aangeboden casus</i>	<i>Tweede aangeboden casus</i>
<i>Arts 1</i>	Casus 1 – App	Casus 2 – EPD
<i>Arts 2</i>	Casus 2 – EPD	Casus 1 – App
<i>Arts 3</i>	Casus 2 – App	Casus 1 – EPD
<i>Arts 4</i>	Casus 1 - EPD	Casus 2 – App
<i>Arts 5</i>	Casus 2 – App	Casus 1 – EPD
<i>Arts 6</i>	Casus 1 – App	Casus 2 - EPD
<i>Arts 7 (test)</i>	Casus 1 – App	
<i>Arts 8 (test)</i>	Casus 1 – EPD	
	<i>Casus 1 = complexe casus (veel ambiguïteit)</i> <i>Casus 2 = eenvoudige casus (weinig ambiguïteit)</i>	

In het experiment kreeg elke arts twee verschillende casussen aangeboden (casus 1 en casus 2), altijd één in de applicatie en één in het EPD. Bij elke arts verschilde dit in de volgorde. De volgorde hiervan is zodanig bepaald dat elke volgorde minimaal twee keer voorkomt, om te kunnen vergelijken of de volgorde (van de casus en van het systeem) van belang in het interpretatie- en besluitvormingsproces.

3.5.4 Omschrijving van de setting

Het experiment werd uitgevoerd in een gereserveerde kamer in het Wilhelmina Kinderziekenhuis in Utrecht. Hierdoor was er zo min mogelijk omgevingsgeluid of afleiding. De respondent werd uitgelegd wat de bedoeling van het onderzoek is. Een uitwerking hiervan is te vinden in bijlage D. Het experiment is opgenomen aan de hand van een schermopname met audio. Zodoende kon bekeken worden wat de arts in het systeem doet op het moment dat een actie of handeling uitspreekt. De respondent werd gestimuleerd om hardop te blijven nadenken. Wanneer het stil viel werd er gevraagd om hardop te blijven nadenken of gevraagd ‘wat denkt u nu?’. De respondent begon bij het overzichtsscherm van het EPD of de app, kreeg daarin een samenvatting van de patiënt en situatie en kon door middel van hyperlinks doorklikken in het systeem. Bij het gebruik van de applicatie kon ook het EPD worden ingezien, aangezien de applicatie geen vervanging is van de huidige systemen en informatie zoals labuitslagen nodig kunnen zijn om tot een diagnose te komen.

3.5.5 Hypotheses

Voor het experiment zijn hypotheses opgesteld ten aanzien van de resultaten. Deze zijn gebaseerd op het theoretisch kader, de verkregen informatie van de ontwikkelaars en projectleiders en de theorie van James March (1987; 2010) over de ambiguïteit van presentatie. Deze hypotheses zijn ook nader toegelicht in het theoretisch kader.

H1: De medicus kan met minder moeite een interpretatie geven aan de data met het gebruik van de app dan met het gebruik van het EPD.

‘Minder moeite’ is geoperationaliseerd met snelheid, de hoeveelheid informatie die bestudeerd werd, de momenten waarop de arts tot een (gedeeltelijke) diagnose komt en momenten waarop de arts het EPD gebruikt. Daarnaast is een vergelijking gemaakt tussen de antwoorden die achteraf bij de follow-upvragen werden gegeven en het gedachteproces.

H2: De medicus komt tot een uitgebreidere differentiaal diagnose met gebruik van de app dan met gebruik van het EPD.

H3: Hoe hoger de mate van ambiguïteit van de casussen, hoe meer voordeel de app geeft voor de interpretatie van de data door de medicus ten opzichte van het EPD.

Met voordeel wordt bedoeld dat de arts met minder moeite en sneller tot een uitgebreidere differentiaal diagnose komt.

3.6 Data-analyse medici

Het audiomateriaal van de think-aloud sessies met follow-up vragen is getranscribeerd en vervolgens geanalyseerd aan de hand van een protocol analyse (Lundgrèn-Laine & Salanterä, 2010). Hierbij is het transcript opgedeeld in syntactische eenheden of secties op één lijn (Fonteyn, Kuipers, & Grobe, 1993, pp.434-435).

Fonteyn, Kuipers en Grobe (1993) beschrijven drie stappen voor de protocol analyse: *referring phrase analysis*, *assertional analysis* en *script analysis* (pp.435-440). Voor dit onderzoek is gebruikgemaakt van de script analyse, omdat deze relevant is voor het interpretatieproces van de medicus. Dit houdt in dat er een beschrijving is gegeven van het redeneringsproces. Er is bekeken op welke informatie de participant focust, op welke manier hij of zij het probleem structureert, de redenering voor de keuzes en de beslissingen die worden genomen en het plan dat ze ontwikkelen voor het oplossen van het probleem. Daarbij werden een aantal *operators* geïdentificeerd om het redeneringsproces te verklaren, bijvoorbeeld observeren/onderzoeken, concluderen, kiezen en verklaren. Dit proces is door Fonteyn, Kuipers en Grobe (1993) verhelderd in een voorbeeld, die ter verklaring ook hieronder is toegevoegd. Uit deze analyse kon worden opgemaakt of de respondenten bij de app of bij het EPD bijvoorbeeld vaker moesten verklaren of bestuderen om tot een besluit te komen. Daarnaast is tijdens de analyse gefocust op de volgende punten:

- Verschillen tussen de twee casussen in het hardop denkproces;
- Verschil in snelheid tussen het gebruik van het EPD en gebruik van de app;
- Verschil in uitgebreidheid van de differentiaal diagnose tussen het gebruik van het EPD en gebruik van de app;
- Verschil tussen de differentiaal diagnoses binnen een casus tussen het gebruik van het EPD en gebruik van de app.

Daarnaast werd er gelet op momenten waarop er (on)duidelijkheid is, er vertrouwen/wantrouwen in het systeem is, de logica van de app wel of niet wordt gesnapt en er een gevoel van zekerheid is over de diagnose. Hierbij is telkens een vergelijking gemaakt tussen het gebruik van het EPD en het gebruik van de app en de mate van ambiguïteit. De antwoorden op de follow-up vragen zijn gecodeerd, eerst open, dan axiaal en ten slotte selectief (Bryman, 2012, p.569). Hiervoor is gebruikgemaakt van NVivo.

TABLE 4: Script Analysis

<i>Data</i>	<i>Operator</i> ¹
<i>Mrs G's heartrate has gradually increased to 130/minute.*</i>	Study ²
I would be on the phone pressing him for alternative medication to break the A-fib sequence	Choose ³
or perhaps suggest setting up for cardioversion.	Explain ⁴
<i>Her respiratory rate has also increased to 32 and she's labored.</i>	Choose
<i>On auscultation, rates are present bilaterally and have increased up to midlung fields.</i>	Study
So, we're not getting any better,	Conclude ⁵
and I would suggest to him that we need to give her more diuretics	Choose
to see if we can't get her to open her kidneys up a bit	Explain
and have some albumin or some colloids at the bedside	Choose
should she lose her pressure.	Explain

*Italics indicate portions of the transcript where the subject is reading from the case study text.

1. *Operator* is defined as a reasoning process.
2. *Study* means to consider information carefully.
3. *Choose* means to decide on an action to take.
4. *Explain* means to provide a rationale for an action.
5. *Conclude* means to decide on the significance, value, or meaning of information.

Naast deze operator analyse, is er onder andere ook gekeken naar het verschil in de hoeveelheid informatie die de artsen observeerden in de applicatie en in het EPD. Het doel hiervan was om te onderzoeken of artsen dezelfde hoeveelheid informatie via de app als via alleen het EPD wisten te verkrijgen. Voor elke casus werd een lijst opgesteld met items, waarvan elk item een stukje informatie representeerde die de artsen uit de app of het EPD zouden kunnen halen. Deze items waren bijvoorbeeld de ademfrequentie of vochtbalans bekijken. Het aantal items per casus dat verworven werd, kon daarmee worden vergeleken tussen artsen die de app gebruikten voor die casus of die alleen het EPD kregen. Deze vergelijking werd gemaakt door eerst de gemiddelde kans dat een item binnen een casus werd waargenomen door de artsen te berekenen voor elke casus, voor elk scenario (app vs. EPD).

3.7 Betrouwbaarheid en validiteit

Twee criteria voor de kwaliteit van het onderzoek zijn betrouwbaarheid en validiteit (Bryman, 2012, p.390). De betrouwbaarheid gaat over de repliceerbaarheid van het onderzoek (p.390). Validiteit houdt in dat er wordt gemeten wat gesteld wordt dat er wordt gemeten (pp.389-390). Omdat het gedeeltelijk kwalitatief onderzoek betreft, zijn deze concepten niet helemaal geschikt omdat er niet van een absolute waarheid kan worden uitgegaan zoals bij kwantitatief onderzoek (Guba & Lincoln, 1994 in Bryman, 2012, p.390). Toch kan er wel worden benadrukt welke maatregelen er zijn getroffen om de betrouwbaarheid en validiteit van het onderzoek zo goed mogelijk te waarborgen.

3.7.1 Betrouwbaarheid

Dit onderzoek betreft een case study en is in die zin niet generaliseerbaar naar andere soortgelijke onderzoeken naar klinische decision-support systemen (Bryman, 2012, p.406). De geïnterviewden gelden niet als representatief voor een grotere populatie (p.406). Dit is ook niet de bedoeling van het onderzoek. De inzichten die zijn verkregen, zeggen iets over het waarde-creatie proces bij deze specifieke decision-support applicatie. Het onderzoek is echter wel volledig transparant met betrekking tot gebruik van de literatuur, dataverzameling en data-analyse, waardoor het wel inzichten kan bieden voor ander onderzoek en ontwikkelingen van decision-support systemen. Ook kan het experiment dienen als blauwdruk voor vervolgonderzoek naar het interpretatieproces van artsen met de komst van nieuwe technologieën die het werkproces en besluitvormingsproces kunnen veranderen.

3.7.2 Validiteit

Door de verschillende technieken voor dataverzameling zijn er ook verschillende validiteitskwesties bij het verzamelen van de data bij ontwikkelaars en het verzamelen van data bij de medici. Deze zullen apart van elkaar hieronder besproken worden.

3.7.2.1 Validiteit van patiëntcasussen

De twee casussen zijn ontwikkeld in samenwerking met artsen die ook betrokken zijn bij de ontwikkeling van de decision-support app, waardoor ze relevant en realistisch waren. De voordelen van een patiëntcasus, een soort simulatie, zijn de relevantie voor de gebruikers (ze testen een situatie die echt voor kan komen), de selectie en bepaling van de taak met de bijbehorende standaardisatie. Omdat het een simulatie is, kon vooraf al bepaald worden hoe het eruit zou komen te zien en konden bepaalde elementen bewust weggelaten of toegevoegd worden. Voor elke medicus waren de casussen hetzelfde, wat de betrouwbaarheid van de resultaten verhoogd (Fonteyn, Kuipers, & Grobe, 1993, p.433). Het nadeel was dat het een incomplete representatie was van een echte situatie; het kan natuurlijk nooit een echte situatie compleet nabootsen (p.434). In een echte situatie zou een arts namelijk een patiënt ook willen zien, wellicht een supervisor spreken en zou hij of zij vaak wel op enige manier bekend zijn met de patiënt.

3.7.2.2 Validiteit van dataverzameling ontwikkelaars

Bij de interviews met de ontwikkelaars zijn de identiteit en de belangen van de onderzoeker kenbaar gemaakt. Dit heeft waarschijnlijk geen probleem gevormd voor de validiteit van de informatie. Het onderzoek zal uiteindelijk ook kunnen fungeren als *usability test* voor de ontwikkelaars en het project UrSmartStatus, dus was het ook in hun belang dat er zo compleet en eerlijk mogelijke informatie werd gegeven.

3.7.2.3 Validiteit van dataverzameling medici

De omgeving en uitleg van het experiment was voor elke participant nagenoeg hetzelfde. De ruimtes wisselden maar waren allen rustig zonder andere aanwezige mensen. Ook vond het voor iedereen plaats

in het Wilhelmina Kinderziekenhuis, waardoor de omgeving vertrouwd was. Alle participanten werkten op dezelfde laptop en kregen dezelfde casussen. De uitleg voor het experiment was gedetailleerd voorbereid dus was ook voor iedereen nagenoeg hetzelfde. Het think-aloud proces ofwel het hardop nadenken ging voor iedereen redelijk soepel af; artsen zijn vaak ook al gewend om aan anderen het gedachteproces uit te leggen waardoor het niet volledig onnatuurlijk voor hen was.

De basis voor de think-aloud techniek en protocol analyse komt van psychologisch onderzoek en een model van informatieverwerking van Newell en Simon (1972). Lundgrèn-Laine en Salanterä (2010) benadrukken dat wetenschappers na Newell en Simon stelden dat verbalisatie een manier is om het cognitieve gedrag van iemand aan te duiden en dat het inzicht geeft in de informatie die in het werkgeheugen opgeslagen ligt op dat moment (p.566). Li et al. (2012) stellen dat het gebruik van cognitieve methodes bij onderzoek naar interacties tussen mens en computer aangemoedigd wordt door actuele best practices (p.762). Het doel van de think-aloud techniek samen met een protocol analyse is om het cognitieve proces, in dit geval het besluitvormingsproces, te verkennen, niet per se om het uiteindelijke besluit te beoordelen (p.566). Dit wordt ook benadrukt door Todd & Benbasat (1987). Volgens hen is protocol analyse een goede methode om vragen rondom het design en het gebruik van een decision-support systeem te onderzoeken (p.493). Het is moeilijk om de invloed van een decision-support systeem te evalueren of meten, omdat het besluitvormingsproces, of zoals Todd en Benbasat het noemen de ‘strategie’ naar een besluit, onnavolgbaar is (Todd & Benbasat, 1987, p.495). Bij het evalueren van het gebruik van een decision-support systeem is het volgens hen gemakkelijker om de kwaliteit van de analyse, het proces, te beoordelen dan de kwaliteit van het uiteindelijke besluit zelf (p.495). Het proces kan namelijk beoordeeld worden aan de hand van normatieve modellen over besluitvorming, het aantal aannames dat wordt gemaakt en in hoeverre er wordt gezocht naar alternatieve mogelijkheden (p.495). Dit zijn benaderingen om de kwaliteit van het besluitvormingsproces te beoordelen. Het besluitvormingsproces is dus een belangrijke factor om de impact van een decision-support systeem te beoordelen (p.495). Een process-tracing methode kan daarom geschikt zijn om de invloed van een decision-support systeem te beoordelen (p.495). Het mentale proces van besluitvorming kan nooit helemaal achterhaald worden, maar met een process-tracing methode kan hier wel zo dicht mogelijk bij gekomen worden.

Uit onderzoek blijkt dat zelfs met een klein aantal participanten de think-aloud techniek veel data oplevert voor analyse (Lundgrèn-Laine & Salanterä, 2010, p.567). Ook is er onderzoek gedaan of het hardop denken het besluitvormingsproces aantast. Daaruit bleek dat het besluitvormingsproces niet veranderd, het kan enkel het proces iets vertragen (Russo, Johnston & Stephans, 1986 in Todd & Benbasat, 1987, p.500; Henry, Lebreck, & Holzemer, 1989 in Fonteyn, Kuipers, & Grobe, 1993, p.432). Er is echter ook kritiek op de methode. Ericsson en Simon (1984) stellen dat het gedachteproces veel sneller kan gaan dan dat het uitgesproken kan worden (Ericsson & Simon, 1984 in Fonteyn, Kuipers, & Grobe, 1993, p.435). Echter zijn onuitgesproken gedachten niet te analyseren, dus door mensen hun gedachten uit te laten spreken en ze aan te sporen dit zoveel mogelijk te doen, wordt er wel zo dicht mogelijk bij het gedachteproces gekomen (Fonteyn, Kuipers, & Grobe, 1993, p.435). Door Nisbett en Wilson (1977) wordt gesteld dat participanten geen goede representatie kunnen geven van hun eigen mentale proces, omdat ze daar geen toegang tot hebben. De auteurs baseren dit echter enkel op onderzoeken met een retrospectieve techniek. Het hardop denken zou ook het gedachteproces van de participant kunnen belemmeren, omdat er maar gelimiteerde capaciteit in het geheugen is (Lundgrèn-Laine & Salanterä, 2010, p.567). Daarnaast kan de onderzoeker te weinig instructies geven waardoor het hardop denken niet op de juiste manier gebeurt (p.567). Ook wordt door sommige wetenschappers de kritiek gegeven dat de techniek zowel observatie als introspectie combineert waardoor er belangrijke informatie gemist kan worden (p.568). Daarbij wordt gesteld dat introspectie de interpretatie van

cognitieve processen door de participant zelf is, en dat deze subjectieve introspectie niet altijd gerelateerd kan worden aan observeerbaar gedrag (p.568). Om dit laatste probleem te voorkomen is het noodzakelijk om als hoofddoel te hebben de subjectieve interpretaties te observeren van de participant (p.568). De think-aloud techniek is niet geschikt wanneer de taak door de participant met automatisme wordt uitgevoerd of als het juist te veel cognitieve spanning vergt (Todd & Benbasat, 1987, p.500). Ook is het niet geschikt als er veel lange termijngeheugen informatie moet worden opgehaald (p.500). Voor dit onderzoek is de think-aloud techniek geschikt, omdat het gedetailleerd inzicht geeft in het gedachteproces, maar zoals aangetoond niet het besluitvormingsproces aantast. De besluitvormingstaak is niet volledig geautomatiseerd maar zal ook niet extreem veel cognitieve spanning vergen, omdat het wel een taak is die de medici dagelijks uitvoeren.

Bij de concurrent techniek wordt het gedachteproces tijdens het besluitvormingsproces uitgesproken, terwijl bij de retrospectieve techniek het gedachteproces uitgelicht wordt na het besluitvormingsproces (Todd & Benbasat, 1987, p.497). De retrospectieve techniek heeft als nadeel dat mensen niet meer precies kunnen vertellen hoe hun besluitvormingsproces ging, zoals ook Nisbett en Wilson (1977) benadrukken. Het voordeel ervan is dat bij het achteraf vragen naar het proces, er niet geïntervenieerd wordt met het proces zelf wat wel het geval is bij de concurrent techniek (p.497). Wat bij de concurrent techniek van belang is om in acht te nemen, is de vraag of de interventie van hardop praten tijdens het besluitvormingsproces een significante impact heeft op het besluitvormingsproces (p.497). Voor dit onderzoek is gebruikgemaakt van de concurrent techniek. Omdat het gaat om soms kleine stappen in het besluitvormingsproces en sommige stappen voor medici automatisch zullen gaan, is het belangrijk om die stappen uit te laten spreken. Wanneer dit achteraf zou gebeuren, is de kans groot dat de participanten bepaalde stappen vergeten of als niet belangrijk achten. In dit onderzoek is het belangrijk dat ook de kleine stappen worden meegenomen in de analyse, omdat dat iets kan zeggen over hoe de informatie uit de app wordt geïnterpreteerd. Ericsson en Simon (1984) stellen dat het bij de concurrent techniek belangrijk is om participanten niet alleen hardop te laten nadenken, maar ook te laten aangeven waarom ze bepaalde stappen zetten. In dit onderzoek is gekozen voor de concurrent techniek, omdat in Todd en Benbasat (1987) wordt aangegeven dat hardop nadenken het besluitvormingsproces niet aantast. De nadelen van de retrospectieve techniek wegen dus zwaarder. Vooral omdat het besluitvormingsproces van de medici gedetailleerd wordt uitgelicht, is de kans groot dat zij achteraf niet precies kunnen reproduceren hoe zij tot bepaalde stappen zijn gekomen.

Todd & Benbasat (1987) maken ook onderscheid tussen gestructureerd en neutraal bevragen (p.498). Bij gestructureerd bevragen worden er specifieke vragen gesteld door de onderzoeker aan de participant over het proces. Bij neutraal bevragen wordt er aan de participant instructie gegeven om het proces te beschrijven, zonder dat de onderzoeker specifieke vragen stelt (p.498). Beide manieren hebben voor- en nadelen. Het voordeel van gestructureerd bevragen is dat het protocol redelijk beknopt blijft en daardoor handig is voor de analyse. Daarnaast is het daardoor ook makkelijker te vergelijken tussen de participanten. Het nadeel is dat participanten op andere informatie moeten focussen, waardoor het proces beïnvloed kan worden. Bij neutraal bevragen is het nadeel dat er minder structuur is en daardoor het categoriseren bij de analyse moeilijker wordt (p.498). Om het besluitvormingsproces zo min mogelijk te beïnvloeden, is er gekozen voor neutraal bevragen. Hierdoor kunnen de medici hun eigen gedachtestappen volgen en hoeven zij niet te focussen op andere informatie.

De protocol analyse heeft als doel om het denkpatroon van de participant te beschrijven en inzicht te krijgen in het besluitvormingsproces (Lundgrèn-Laine & Salanterä, 2010, p.570). Voor de protocol analyse worden er ook voor- en nadelen genoemd. Voordelen zijn dat het gedetailleerd kan weergeven op welke informatie de participant zich focust tijdens het proces en de strategieën die de participant

hanteert om een besluit te maken of oplossing te vinden (pp.567-568). Protocol analyse is minder geschikt wanneer het wordt gebruikt voor de analyse van het gedachteproces bij slecht gestructureerde, grote taken in complexe omgevingen (p.568). Van dat laatste is geen sprake.

4. Resultaten ontwikkelaars

Om de empirische deelvraag *Hoe wordt waarde toegekend aan medische data door de ontwikkelaars van UrSmartStatus?* te beantwoorden, zijn er vijf interviews gehouden met de ontwikkelaars waarvan twee interviews met twee respondenten tegelijkertijd. Deze interviews zijn gecodeerd met behulp van NVivo waarbij een codeboom is opgesteld, te vinden in bijlage C. Hieruit kwamen drie thema's naar voren, namelijk het verschil tussen de app en het EPD, besluitvorming en interpretatie van data. Deze thema's zijn elk ook weer onderverdeeld in subthema's. De resultaten zullen worden besproken aan de hand van de drie thema's waarin ook de subthema's aan bod komen.

4.1 Verschillen tussen de applicatie en het EPD

Een deel van de waarde-creatie wordt duidelijk bij het duiden van het verschil tussen de app en het EPD. Deze verschillen (en overeenkomsten) worden letterlijk genoemd, maar de duiding hiervan kan ook gevonden worden in de beschrijving van de functie of het doel van de app. Hoe in het proces keuzes worden gemaakt en hoe en of er verantwoording over wordt afgelegd, wordt ook beschreven.

Over het verschil tussen de app en het EPD wordt onder andere gezegd dat de app zal leiden tot snellere beslissingen of beeldvorming, dat het intuïtiever en handiger is voor de arts (R1, R2). Dit zou komen doordat het beter aansluit bij hoe artsen denken, waardoor het intuïtiever is. Dit idee sluit aan bij het idee van March (2010) over het ambigu presenteren van informatie. Het ambigu presenteren van informatie sluit namelijk beter aan bij hoe mensen denken. Op die manier zien de ontwikkelaars dat zij waarde toevoegen aan de data:

“Ik zou dat willen samenvatten in ‘intuïtief’, dat dat het grootste verschil gaat worden [met het EPD]. Want artsen denken niet na in het tabblad. [...] Die denken in ‘tabblad hoe gaat het met het hart’, ‘hoe gaat het met de longen’, ‘hoe gaat het met de nieren’ en die is er niet. En niet in een volledig arbitraire verdeling die zo gekozen is, omdat het gemakkelijk is voor de programmeur om te implementeren, maar omdat het logisch aansluit op z'n werkprocessen en denkwijze” (R1).

Het huidige systeem is zo opgezet dat de arts zelf veel moet interpreteren en stukjes data met elkaar moet verbinden, terwijl met de app dit al voor een deel wordt gedaan. Dit wordt door een respondent ook als een verbetering gezien (R4). Dat het intuïtiever is, heeft ook te maken met dat de applicatie het werkproces van de arts centraal stelt (R7). De manier van denken en werken wordt als uitgangspunt genomen. Dit is een verschil met het EPD, waarin het meer pragmatisch is ingesteld. Een ander verschil met het EPD is de conclusie die de app verbindt aan de data, in de vorm van de kleuren van de organen (R3). Dit wordt ook wel benoemd als dat de applicatie een overweging meegeeft en de data spiegelt aan wat normaal is voor vergelijkbare patiënten (R7). Op die manier zegt het iets over de data, terwijl het EPD dat niet doet. Dit komt door de hoeveelheid data die de app verwerkt. De data van historische patiënten wordt namelijk gebruikt voor de data die het algoritme analyseert: *“[...] het is natuurlijk wel heel erg anders omdat er een gigantische hoeveelheid meer informatie achter zit” (R1).* Zoals ook in het theoretisch kader besproken, vraagt deze grote hoeveelheid aan data om zorgvuldigheid van de verwerking ervan. Dat de respondenten inzien dat deze hoeveelheid aan data een groot verschil is met hoe de systemen nu werken, laat zien dat ze bewust zijn van de impact die het kan hebben.

De toegevoegde waarde van de app ten opzichte van het EPD is dus dat het ervoor zal zorgen dat de arts sneller en intuïtiever een overzicht heeft; het werkproces van de arts wordt centraal gesteld. Een deel van de interpretatie wordt al gedaan waarmee conclusies worden verbonden aan de data middels het vergelijken met historische patiënten. Dit wordt mogelijk gemaakt door de grote hoeveelheid data die erachter zit en de algoritmes. Toch wordt door een deel van de respondenten de app benoemd dat het inzetten van een dergelijke app niet veel anders hoeft te zijn dan het gebruiken van andere hulpmiddelen.

Uit het theoretisch kader bleek echter uit bijvoorbeeld het artikel van Mittelstadt et al. (2016) dat het inzetten van zulke grote hoeveelheden data implicaties kan hebben voor hoe er met de data kan worden omgegaan. Het verbinden van conclusies aan de data zorgt ervoor dat de ontwikkelaars een deel van het interpretatieproces van de arts overnemen, dus het is de vraag of dit goed gebeurt, hoe inzichtelijk dit voor de arts is en hoe erover gecommuniceerd wordt.

4.1.1 Het doel van de applicatie

Uit de interviews komt naar voren dat de respondenten het grotendeels met elkaar eens zijn over wat het uiteindelijke doel van de app zou moeten zijn. Door bijna alle ontwikkelaars wordt in verschillende bewoordingen gesteld dat de app een visuele samenvatting geeft van de verschillende relevante parameters; een overzicht van hoe het met de patiënt gaat (R1, R2, R4, R5, R6, R7). Daarmee wordt het verhaal van de patiënt verteld (R2). Dit overzicht of deze samenvatting wordt omschreven zoals March (1987) het ambigu presenteren van informatie omschrijft:

“Juist doordat je eigenlijk maar een aantal stappen in het informatieproces geeft, dus de informatie een beetje grof samenvat, vertel je misschien wel beter het verhaal dan dat je elk ding voor iemand uit gaat kauwen. Waarbij die eigenlijk na stap 3 al afhaakt en dan eigenlijk cruciale stap 6 even had gemist. Dat is eigenlijk een beetje hoe dat samen zou moeten werken. Hoe dat op die manier zou kunnen gebeuren. Of dat zo is dat is iets wat nog moet blijken. Dat is wel een beetje wat de achtergrond daarvan is” (R7).

Hiermee stelt deze respondent dat de arts dus niet voorgeschoteld krijgt wat hij er precies van moet vinden, maar een samenvatting krijgt die genoeg overlaat aan de eigen (nodige) interpretatie van de arts. De vergelijking wordt gemaakt met een film, waarin niet alle letterlijke gebeurtenissen getoond worden, maar door de juiste gebeurtenissen wel te tonen, vult ons brein in wat er gebeurd is. Het hele verhaal wordt vanzelf door het brein ingevuld. Dit wordt door de respondent ook wel het ‘faciliteren van de interpretatie’ genoemd (R7). Met de vergelijking van de interpretaties van de artsen zal blijken op welke manier hun interpretatie gefaciliteerd wordt en of er genoeg interpretatie wordt overgelaten. Er wordt namelijk wel degelijk al een interpretatie aan de data gegeven door met de kleuren de ernst aan te geven. Het is de vraag of de artsen dit net zo interpreteren (Mittelstadt et al., 2016, p.5) en de rationele logica kunnen volgen en hiermee omgaan (Wang & Ruhe, 2007; Peterson, 2009). Door het overzicht wordt het snel duidelijk waar de aandacht van de arts naartoe moet. Het geeft een prioritering aan de data, zo stellen ook bijna alle respondenten:

“Wat ik denk dat onze app vooral doet eigenlijk verschillende databronnen met elkaar combineren en ook gelijk filteren en bijna proto-interpretatie geven van wat erachter ligt om zo een arts of wie dan ook al een richting op te sturen van goh, let hier op, dit is belangrijk en dit niet” (R6).

Met de prioritering die wordt gegeven wordt de arts dus ook een bepaalde richting opgestuurd naar waar zijn aandacht moet liggen, aldus respondent 6. Uit het theoretisch kader bleek dat het de vraag is of dit wenselijk is. Hiervoor zou het algoritme heel accuraat moeten zijn, waarbij het ook de vraag is of het dat kan zijn door de ambiguïteit van medische data en de noodzakelijkheid van ervaringskennis van artsen.

Naast dat de arts een overzicht krijgt van hoe het met de patiënt gaat, zijn de ontwikkelaars ook bezig met het creëren van een functie waarmee de gezondheid van de patiënt in de toekomst wordt voorspeld. Het uiteindelijke doel van de app is om de arts betere beslissingen te laten nemen, bijvoorbeeld doordat de app voorspelt dat het straks goed gaat met de patiënt en er minder zal worden overbehandeld. Deze prognose wordt gebaseerd op historische patiëntgevallen. Het doen van een voorspelling is wat de arts nu zelf doet, dus de app neemt ook dat gedeeltelijk over. Ook hierbij geldt bovenstaande vraag hoe er dan wordt omgegaan met ambiguïteit en onzekerheid van data en hoe de ervaringskennis van artsen daarin een rol speelt.

Ten slotte wordt er door drie van de ontwikkelaars benadrukt dat de app geen diagnosetool is, maar meer gezien moet worden als een monitoringssysteem of alarmsysteem. Het is daarmee een ondersteuning voor de arts en vormt niet de basis waarop het besluit wordt genomen (R4). Toch is het in deze fase van de ontwikkeling nog lastig te zeggen hoe het uiteindelijk zal worden gebruikt: *“Dit is zo’n veelzijdig ding dat je er honderd verschillende usecases mee kan verzinnen en het is net wie je vraagt welke het belangrijkste zijn”* (R6). De functies zullen ook afhangen van het doel van de app: *“Gebruik je onze applicatie om iets te zeggen over de individuele patiënt of is het meer iets om te discrimineren van waar besteed ik nu m’n tijd als eerste aan?”* (R7). Het doel van de app heeft gevolgen voor de impact die de app kan hebben. Dat het doel nog onbepaald is, maakt de impact die het zal hebben ook nog onbepaald. Dit kan voor de ontwikkelaars wellicht ook lastig zijn, omdat zij niet weten hoe hun app straks gebruikt zal worden. Het zorgvuldig omgaan met de data kan daardoor een andere lading krijgen, omdat niet precies duidelijk is wat er mee zal gebeuren.

De app geeft dus een visuele samenvatting met prioritering van de data waarbij er genoeg ruimte wordt overgelaten voor de interpretatie van de arts, zo wordt gesteld. De vraag hierbij is hoe de interpretatie van de arts gefaciliteerd wordt. De arts wordt wel in een bepaalde richting gestuurd, maar heeft wel ruimte voor eigen interpretatie. Hoe verhoudt zich dit tot elkaar en hoe gaan artsen hier mee om?

4.1.2 Keuzes in de applicatie

Bij het ontwikkelen van de app worden door de ontwikkelaars keuzes gemaakt. Deze worden vaak in overleg met elkaar gedaan, waarin er ook veel discussie en verdeeldheid kan zijn. Het is niet altijd duidelijk of iets de juiste keuze is, soms kan iets ook een technische keuze zijn. Soms is het een combinatie van wat de gebruiker zou willen en wat er technisch mogelijk is in de tijd. Wat ook vaak gebeurt, is dat er snel een beslissing voor iets wordt gemaakt en dat dit later wordt getoetst of gevalideerd bij de andere ontwikkelaars of de betrokken artsen, een soort *trial and error*. Over specifieke beslissingen wordt ook gesproken. Zo worden beslissingen met medische betekenis altijd in overleg gedaan met de betrokken artsen bij het project, maar worden beslissingen voor de codes zelf gemaakt. Het doen van gebruikerstesten is in deze fase nog moeilijk. Wanneer de ontwikkeling van de app wat verder is gevorderd, zullen wel artsen en patiënten bevroegd worden.

Er werd verschillend gereageerd op de vraag of er (wetenschappelijke) literatuur werd gebruikt. Sommige ontwikkelaars gebruiken dit wel: *“Er is wel literatuur bij gehaald, vooral met betrekking tot grafieken en dergelijke en hoe geef je nou visueel goed dingen weer”* (R5) maar sommige ook niet: *“Wat dat betreft is er niet echt literatuuronderzoek gedaan, niet door mij in ieder geval [...]”* (R3). Respondent 7 geeft aan dat er, waar mogelijk, wel wetenschappelijke achtergrond wordt gebruikt. Dit lijkt erop dat de mensen die het project managen wel bepaalde beslissingen maken die voortkomen uit literatuur, maar dat er in de praktische ontwikkeling ook keuzes worden gemaakt die pragmatisch zijn.

Over de verantwoording van de keuzes wordt door respondent 1 en 2 aangegeven dat er per week of per project wordt bijgehouden wat de agendapunten zijn, wat er gedaan gaat worden en wat de genomen besluiten zijn. Iedere week is er een vast overleg waarbij *“[...] in verschillende mate van trouwheid wordt genotuleerd”* (R1). Respondent 5 en 6 geven echter aan dat er op dit moment geen verantwoording wordt afgelegd, dus zien het wekelijkse overleg niet per se als een manier van verantwoording. Zij zien wel de noodzaak ervan wanneer het project zich verder vordert:

“Nadat we de ontwikkelingsfase ingaan, dat is het hele certificeringstraject. Dus wat we doen is zo goed mogelijk onze eigen dingen documenteren. En dan echt zodra we het certificeringstraject

ingaan en daadwerkelijk testen willen gaan uitvoeren, het zo goed mogelijk proberen te matchen met wat we eigenlijk nodig hebben. En dan zo goed mogelijk verantwoording afleggen” (R6).

Een certificeringstraject zal uiteindelijk ook voor zorgen dat er geen dingen worden geïmplementeerd die het werk van de arts negatief zullen beïnvloeden en er gegronde redenen zijn waarom de app een toegevoegde waarde zal zijn.

Het lijkt erop dat het gebruik van literatuur vooral bij de projectleiders bekend is en op basis daarvan de app richting krijgt. Besluiten binnen deze richting gebeuren in overleg en middels trial and error. Voor het certificeringstraject zal ook noodzakelijk zijn dat keuzes onderbouwd worden en testfasen zullen uitsluiten wat al dan niet werkt.

4.2 Besluitvorming: het verschil tussen de logica van de arts en de applicatie

De waarde-creatie van de ontwikkelaars in de app kan ook geduid worden door te kijken naar het verschil tussen hoe een arts nu een besluit maakt en hoe de app dit kan veranderen of gedeeltelijk overneemt. Hoe de app daarin communiceert met de arts en hoe de arts omgaat of om zou moeten gaan met de app speelt daarin een rol. Tevens is het belangrijk om te bekijken welke rol het vertrouwen van de arts in de app speelt, omdat dit ook kan bepalen hoe de arts met de app omgaat en welke waarde hij hecht aan de betekenis die aan de data wordt gegeven.

Een duidelijk verschil dat door vier respondenten wordt benoemd, is het vergelijk dat de app kan maken tussen de huidige patiënt met heel veel historische patiënten, waarbij een arts zijn ervaring kan gebruiken maar nooit met zoveel precisie en grootte van gegevens kan vergelijken:

“[...] voor een gedeelte is het ook bundelen van ervaringen uit het verleden. [...]om puur op een menselijk cognitieve capaciteit zeg maar die vergelijking te maken van oké, welke patiënt heb ik gehad, lijkt die misschien op die. En dan hebben we het over verschillende parameters van hart tot long tot nieren [...]. Moet je dat allemaal een soort van in de tijd vergelijken. Dat is eigenlijk een onmogelijke opgave, terwijl een computer dat” (R2). “Dat zijn juist de dingen die een computer goed kan” (R1).

Het doel daarvan is dat het leidt tot nauwkeurigere diagnoses en beleid van artsen, omdat de algoritmes *“getraind zijn op duizenden en duizenden patiënten, wat hopelijk een accuratere uitkomst geeft dan een enkele arts” (R4).*² Het vergelijken met historische patiënten is iets wat een arts nu in principe ook doet, alleen kan hij dat inderdaad niet met zoveel precisie en ook niet met zoveel patiënten. Daarnaast zou het wellicht het probleem van omgaan met onzekerheid en kansen van artsen gedeeltelijk kunnen oplossen (Berger, Bleichrodt & Eeckhoudt, 2013, p.559; Camerer & Weber, 1992, p.325; Carney et al., 2007, p.234; Kahneman & Tversky, 1972). Met de ‘objectieve’ vergelijking die de algoritmes maken met historische patiënten, is de voorspelling niet meer afhankelijk van hoe een arts met onzekerheid en kansen omgaat, maar worden de principes van waarschijnlijkheid toegepast. Intuïtie en subjectiviteit over de situatie zal dan minder een rol spelen. Uiteraard is wel de vraag hoe dit tot elkaar zal verhouden; ervaring en intuïtie zal altijd noodzakelijk blijven omdat het om individuele patiënten gaat. Er zal dus een juiste verhouding moeten worden gevonden tussen het oplossen van de negatieve effecten van de subjectiviteit van de arts en het in stand houden van de noodzakelijkheid ervan.

De app maakt al een soort suggestie van welke data bij elkaar hoort en bekeken moet worden. Op die manier heeft het invloed op de besluitvorming van de arts, omdat hij al een bepaalde richting wordt opgestuurd. Dit wordt door vier respondenten genoemd. Hierbij wordt de vergelijking met de Gestalt psychologie gemaakt door een respondent, waarin losse data gegroepeerd wordt waardoor de associatie

² De quotes van respondent 4 zijn vertaald uit het Engels.

wordt gemaakt dat bepaalde data bij elkaar hoort of verband met elkaar heeft. Dit wordt straks door het algoritme gedaan, maar op dit moment moet de arts die verbanden zelf leggen:

“Door al een suggestie te doen van we verwachten dat deze patiënten bij elkaar horen dus daar teken je dan spreekwoordelijk zo ’n hokje omheen van dit is één eenheid. Terwijl op dit moment in HiX en Metavision heb je eigenlijk niet die hokjes die getekend worden, daar heb je gewoon de ruwe, feitelijke data zoals die op dit moment in de database zit” (R1). “Het hokje moet dan zelf gemaakt worden door de arts” (R2).

De app presenteert dus gewaardeerde data aan de arts waardoor de arts in een bepaalde richting gaat denken. Uit de accuraatheid ervan zal moeten blijken hoe kritisch de arts hier tegenover moet staan om geen andere ‘hokjes’ te missen die wellicht ook van toepassing kunnen zijn. De richting die de arts wordt opgestuurd moet of accuraat genoeg zijn, of uitnodigen om die richting te controleren.

Daarnaast heeft de app een overzicht van hoe het met de patiënt gaat, op een manier dat een arts niet kan. De app neemt elke waarde van een orgaan mee in het beeld van de patiënt, terwijl een arts nooit op die manier het overzicht kan hebben. Het menselijk brein kan niet tegelijkertijd overal de aandacht op vestigen (R4). In die zin kan de app wel een beter beeld geven van de gezondheid van de patiënt. Andersom is er ook een belangrijk verschil, namelijk dat het algoritmisch model vooraf bepaalde parameters meeneemt, terwijl een arts ook een parameter kan opmerken die een belangrijk effect kan hebben maar niet in het model wordt meegenomen (R6). Bijvoorbeeld in het geval van speciale casussen, waarbij het model zoiets nog nooit gezien heeft en er dus ook niets over kan zeggen. Terwijl een arts er wel kennis over kan hebben vanuit de studie, artikelen of ervaringen van andere artsen (R5).

Hoewel het gebruiken van de app als anders wordt gezien dan hoe een arts nu een besluit maakt, zijn er ook processen die ook al op deze manier verlopen. Daarbij wordt het voorbeeld genoemd van de echografist die de echo’s interpreteert en dit deelt met de arts. Op die manier krijgt de arts *“iets wat gecategoriseerd is door een black-box”* (R1). Dit proces doet zich in de huidige situatie voor en de vraag is of dit anders is dan wat het algoritme straks zal doen. Ook wordt gezegd dat er in beide gevallen een besluit wordt genomen op basis van een model. Op dit moment op basis van het model in het hoofd van de arts, *“waarbij hij kijkt van joh, ik weeg wat dingen af, ik zie wat dingen, ik kijk ernaar, de patiënt is zo, dit zal waarschijnlijk de beste behandeling of de juiste diagnose zijn”* (R5). In de app zal het model gebaseerd zijn op veel meer patiënten, enkel is het algoritme niet bevooroordeeld *“[...] want als je als arts de vorige keer een patiënt gemist hebt, dan ben je de volgende zes keer voorzichtiger”* (R5). De app maakt bij elke patiënt dezelfde afweging, ongeacht de uitkomst van vorige patiënten. De app zou daarin objectiever zijn dan de arts, omdat het bij elke patiënt een *“exacte kansberekening”* is (R5). De theorieën over het omgaan met onzekerheid en kansen zijn dus bekend bij de ontwikkelaars, zoals hierboven beschreven. De *bias* die een mens heeft door bijvoorbeeld recente patiëntgevallen die beter in het geheugen zitten of het hebben van een slechte dag, wordt door de app niet meegenomen (R7). In dat opzicht is de app objectiever dan de mens. Aan de andere kant kan de applicatie de ambiguïteit van die statistische schatting van de data niet meenemen, daar is de mens juist weer voor nodig. Het is dus de vraag of de objectiviteit van de algoritmes kunnen werken met de ambiguïteit van de medische data.

Drie van de respondenten geven aan dat de app het besluitvormingsproces zal verbeteren door *“de computer de dingen te laten doen die de computer goed kan en de mens de dingen te laten doen die de mens goed kan”* (R1). Een van de respondenten stelt daarbij ook dat het niet veel zal veranderen, enkel het proces aangenamer en efficiënter maken en in de toekomst accurater (R4). Dat is opvallend, omdat uit het theoretisch kader bleek dat het besluitvormingsproces wel degelijk zal veranderen door de waarde-creatie van de ontwikkelaars. De arts krijgt een soort suggestie en richting en moet daar zelf kritisch tegenover blijven staan, waarbij hij anders zelf de ruwe data interpreteert. Deze ontwikkelaars

zijn wellicht optimistisch over de werking van de app en het oplossen van de problemen zoals ambiguïteit van medische data. Ze zijn zich ervan bewust dat de arts op een andere manier gaat werken, maar zijn wellicht van mening dat het uiteindelijke besluit niet veel anders zal zijn. Een andere respondent vraagt zich af of het zal leiden tot een betere diagnose, maar hoopt wel dat het ertoe zal leiden dat artsen minder snel iets over het hoofd zien (R7).

De applicatie maakt een vergelijking met een grote dataset aan historische patiënten, iets wat een arts niet kan doen. Mensen hebben namelijk een beperkte capaciteit om geschiedenis in het geheugen te bewaren en zijn gevoelig voor herinneringen die passen bij hun huidige overtuigingen (March, 2010, p.45). Aan de ene kant maakt dat de vergelijking objectiever; ervaringen met kortgeleden patiënten of emoties worden niet in de overweging meegenomen. Aan de andere kant mist de vergelijking de intuïtie en klinische kennis die een arts wel heeft, wat de beoordeling van een individuele patiënt nodig heeft. Het overzicht dat de app kan maken van de gezondheid van de patiënt is iets waar een arts veel meer moeite voor moet doen, als hij het überhaupt al op een zodanige manier kan krijgen. De applicatie kan veel gemakkelijker alle data met elkaar combineren, terwijl het menselijk brein die capaciteit niet heeft. Het overzicht kan er wel voor zorgen dat artsen minder snel iets over het hoofd zien. De app doet daarbij een suggestie over de gezondheid van de patiënt. Hoe moet de arts hier mee omgaan?

4.2.1 Communicatie tussen de arts en de applicatie

Hoe de app communiceert of zou moeten communiceren met de arts en hoe de arts omgaat of zou moeten omgaan met de app kan van invloed zijn op de besluitvorming van de arts. De manier waarop ontwikkelaars kijken naar de manier waarop de app zou moeten communiceren en artsen om zouden moeten gaan met de app, kan iets zeggen over op welke manier de app past binnen het besluitvormingsproces van de arts. De inzichtelijkheid of transparantie van de data en analyse en het vertrouwen van artsen in de app kan daarin een rol spelen.

Ten eerste wordt door drie respondenten genoemd dat de app moet communiceren op een manier die aansluit op hoe mensen denken. Dus geen kansen weergeven, want mensen kunnen daar niet goed mee omgaan maar *“dat je de nummers vertaalt naar de beleavingswereld van je personeel, zonder dat je daadwerkelijk iets verandert”* (R1). Concreet betekent dit voor de voorspelling van de gezondheid van de patiënt dat er geen kansen, percentages of cijfers worden gegeven, maar dat de geschiedenis van de patiënten die op de patiënt lijken wordt weergegeven. Op die manier wordt er geen claim gemaakt met percentages maar is het enkel een vergelijking met patiënten die op de huidige patiënt lijken. Hierbij wordt ook aangegeven dat de applicatie een communicatieprobleem tussen de computer en de mens oplost, het probeert de twee systemen elkaar beter te laten begrijpen door informatie op een betere manier te presenteren (R7). Dit komt overeen met de theorie van James March (2010): hoe meer ambigu iets gepresenteerd wordt, hoe beter het past bij hoe mensen denken. De app geeft dus geen keiharde uitspraken, maar de arts moet zelf blijven nadenken over welke conclusies hij daaraan verbindt en daar kritisch over reflecteren (Schön, 1991). Daarbij blijft de arts altijd eindverantwoordelijke (R5). Deze respondent stelt dat het gebruiken van de app in principe niet veel anders is dan nu, dat de arts nu ook een soort model in zijn hoofd heeft op basis waarvan hij keuzes maakt, maar straks de app gebruikt die enkel een grotere dataset heeft en minder bevooroordeeld is. Dit sluit aan bij wat andere respondenten zeggen over hoe artsen nog steeds de bewuste keuzes maken en de app daarin een hulpmiddel blijft. De beslissing wordt een samenspraak van de arts en de app. De app doet een suggestie, maar de arts maakt daarin keuzes wat hij wel en niet wil gebruiken en welke data erin blijft staan en eruit wordt gehaald:

“Eigenlijk hoop je met deze applicatie dat je juist meer een gesprek maakt tussen die data en die persoon, zodat die persoon misschien niet alleen in die applicatie maar ook in z'n eigen gedachteproces er alsnog die ambiguïteit bij kan voegen. Of tenminste, niet ambiguïteit maar

dat het begrip van de situatie aan kan toevoegen waarmee een deel van die ambiguïteit nog opgelost kan worden” (R7).

In een definitieve versie zou het mogelijk moeten zijn om als arts aanpassingen te maken (R7). De app maakt geen autonome keuzes. Daarbij is het de bedoeling dat de arts niet ‘voorgescheteld’ krijgt wat hij ervan moet vinden, maar een samenvatting krijgt waarbij hij zelf nog interpretatie aan moet toevoegen. Het laat zien wat er normaal gebeurt en wat de consequentie daarvan is, maar suggereert geen interventies en communiceert zo open mogelijk over wat er gebeurt (R2, R6, R7).

Uit bovenstaande blijkt dat de ontwikkelaars de inspraak van artsen belangrijk vinden. De ontwikkelaars voegen waarde toe in de vorm van een samenvatting en ernst-interpretatie, maar de arts moet daar wel kritisch op reflecteren want hij blijft eindverantwoordelijke. De rol voor het kritisch inzicht van de arts blijft dus bestaan.

Over de omgang met de app wordt gezegd dat dit waarschijnlijk erg gaat afhangen van de individuele arts, de één is kritisch en de ander neemt sneller iets voor waar aan:

“[...] hoe gaan verschillende mensen met uiteindelijk wel hetzelfde over-all profiel, maar met verschillende niveaus van ervaring, om met de applicatie. Want het moet niet zo zijn dat alle jonge artsen alles klakkeloos overnemen, terwijl alle senior artsen eigenlijk maar volledig negeren en toch wel een eigen plan trekken, want dan gaat er ergens iets fout.” (R1)

Hier wordt aan toegevoegd dat uiteindelijk de arts feedback zou moeten kunnen geven in de app, dat hij kleuren of waardes kan aanpassen als hij het idee heeft dat de app het fout heeft (R5). Wanneer data op een bepaalde manier gepresenteerd wordt, zou er rekening moeten worden gehouden met hoe de artsen dit interpreteren. Als de een erg kritisch is en de ander alles klakkeloos overneemt, wordt er op verschillende manieren waarde gegeven aan de app en de data die wordt gepresenteerd.

De app geeft een soort conclusie aan de data door de organen rood, oranje, geel of groen te kleuren, zo wordt gesteld (R3). Toch moet de arts blijven interpreteren, en doorklikken om te zien wat er achter zit. Op die manier worden eventuele fouten of niet-werkende apparatuur ondervangen. Toch zit daar wel meer risico in dan hoe dit nu met HiX en Metavision gaat:

“Mits het slecht gaat, maar als de app zegt het gaat hartstikke goed, en het blijkt toch niet goed te zijn, ja dan heb je wel een probleem. Want dan ben je geneigd oké alles is groen, oké door. Ja dat is een goeie vraag, maar ja dan denk ik ook je moet ook altijd weer terug naar die patiënt van hoe ligt die erbij, en gaat het goed?” (R3).

De kleuren moeten “vooral uitnodigen om verder te zoeken” (R6). Dat er niet direct conclusies uit worden getrokken, maar dat het een signaal geeft om uit te zoeken wat er achter zit. De arts kan wel sneller inzien wat belangrijk is en wat niet, omdat de kleuren een prioritering aanduidt van de data. Dit wordt ook benoemd als de kwaliteit van data: “De kwaliteit van informatie die ook gecommuniceerd moet worden” (R7). De aangegeven kleuren zeggen daar iets over. Toch wordt ook ingezien dat er wel situaties kunnen zijn die problematisch kunnen zijn. Als ontwikkelaars bepalen zij namelijk van tevoren welke parameters worden meegenomen in de analyse, terwijl een arts een bepaalde parameter als belangrijk kan achten die niet wordt meegenomen, maar wel veel effect kan hebben. De respondent zegt hierover: “[...] op het moment dat ze uit elkaar liggen, dan zou ik eerder de arts vertrouwen dan het model” (R6) In die zin blijft de arts verantwoordelijk voor het maken van bewuste keuzes en het kritisch beoordelen. Deze opvattingen over de kleuren zijn opvallend. De ontwikkelaars geven waarde aan de data door ze te verbinden met kleuren, maar de arts moet wel achterhalen wat er achter die kleuren zit. De kleuren hebben dus nog wel de interpretatie van de arts nodig, terwijl de ontwikkelaars met de kleuren ook al interpretatie hebben gegeven aan de data. De vraag die daarbij opkomt is welke waarde

de kleuren dan hebben voor het besluitvormingsproces. Hoe wordt er gecommuniceerd dat de kleuren ‘uitnodigen om verder te zoeken’?

Uit bovenstaande blijkt dat de interpretatie van de arts erg belangrijk blijft. Onduidelijk is wel hoe de arts hier mee om gaat: hoe wordt duidelijk dat de eigen interpretatie van belang is om eventuele fouten te onderscheppen? De arts blijft eindverantwoordelijke, maar moet ook af kunnen gaan op de kleuren om vertrouwen in het systeem te hebben. Hoe verloopt dit in situaties dat organen groen kleuren, maar eigenlijk rood hadden moeten zijn?

4.2.2 Inzichtelijkheid van data en analyse

Er is de ambitie om inzichtelijk te maken waar de kleuren op gebaseerd zijn, wat de grenzen zijn die zijn ingesteld (R1). Hier wordt ook op gezegd dat je niet alles wil laten zien, en dat het de vraag is hoeveel ‘lagen’ je verder wil gaan (R2). Er moet een middenweg worden gevonden waarin het duidelijk wordt wat heeft geleid tot de categorisering, maar het niet alsnog een overload aan informatie wordt: *“De vraag is hoeveel nadruk het moet hebben, want als onze applicatie goed werkt, zou het juist intuïtiever moeten zijn en zou dat allemaal niet per se nodig hoeven wezen. Wat niet wil zeggen dat je dan niet af en toe denkt van hé, maar hoe zit dit?”* (R7). Er moet in ieder geval gecommuniceerd worden dat het gebaseerd is op vergelijkbare historische patiënten en dat de voorspelling een voorspelling is, geen situatie die sowieso gaat gebeuren (R3). Het is wel volgens deze respondent de vraag in hoeverre je uitleg moet geven over het model:

“Ik denk dat je altijd een uitleg moet geven als er vragen zijn, maar of de verpleegkundigen daar echt op zitten te wachten dat die echt een soort uitleg van dat model, ik weet niet of ze daar echt waarde aan hechten, en of ze het echt zullen begrijpen.” (R3).

Respondent 4 stelt ook dat als de artsen kunnen zien waarom een bepaalde voorspelling wordt gedaan, dat het een meerwaarde zou zijn voor de app. Het is wel de vraag hoeveel details een arts nodig heeft of wil hebben. Op het moment dat er niet genoeg data is om iets te kunnen zeggen over een patiënt, zal de app dit ook moeten communiceren: *“Het is altijd goed om zo transparant mogelijk te zijn. Het is nog uitzoeken hoe je het het beste kan communiceren. Maar het streven is wel om open en transparant te zijn”* (R6). Er wordt ook gezegd dat het juist niet volledig transparant is. De arts kan de waardes inzien, maar hij of zij kan niet zien dat bij waarde X het orgaan rood wordt en bij waarde Y het orgaan groen:

“Dat is wel iets wat indirect de artsen de vrijheid geeft om zelf die cut-off points te bepalen. De stip op de horizon is dat het wel enigszins inzichtelijk is van oké, we hebben deze grenzen gesteld dus als hartslag die en die waarde is, zou dat rood of groen of iets moeten betekenen” (R1).

De uitleg van het model wordt door een respondent gezien als een ‘nice to have’ (R3). Respondent 6 is hierin stilliger en stelt dat het goed is om zo transparant mogelijk te zijn. Ook deze respondent geeft aan dat het nog niet duidelijk is hoe dit het beste gedaan kan worden.

De communicatie van de app is dus nog een vraagstuk. Zowel de manier waarop de arts er mee om moet gaan (de kleuren moeten ‘uitnodigen om verder te zoeken’, maar hoe wordt dit gedaan?) als de mate van transparantie van de data en analyse is niet duidelijk. Dit heeft echter wel consequenties voor de impact die het heeft op het besluitvormingsproces van de arts. De arts moet er mee zien om te gaan, maar op welke manier is nog niet duidelijk. Dit zal voor de interpretatie van de arts wel belangrijk zijn. Wie is er verantwoordelijk wanneer er foutieve correlaties worden gegeven?

4.2.3 Vertrouwen van de arts in de applicatie

Over het vertrouwen dat de arts in de app moet hebben, worden verschillende dingen gezegd. Zo wordt aangegeven dat de app vergelijkbaar is met alle andere hulpmiddelen die artsen gebruiken, en dat het vertrouwen daarin op dezelfde manier kan worden beschouwd: *“Is het goed of slecht dat een arts*

vertrouwen heeft in een radiologierapport?” (R1). Ook wordt aangegeven dat er zowel vertrouwen moet zijn in het systeem wanneer het goed getest en gevalideerd is, maar dat artsen ook zelf moeten blijven nadenken (R3). Daarmee dient de app als ondersteuning voor de arts. Als aanvulling daarop stelt deze respondent dat dit nu ook het geval is. Wanneer er bijvoorbeeld fouten in het systeem zitten of apparatuur niet goed gekalibreerd is, moet de arts dat ook herkennen en interpreteren. Het verschil met de app is dat er al een conclusie wordt getrokken uit die data. Wanneer het slecht gaat met de patiënt zal de arts daarop doorklikken en bekijken wat de waardes zijn, maar in het geval het goed gaat kan dit wel een probleem zijn:

“[...] maar als de app zegt het gaat hartstikke goed, en het blijkt toch niet goed te zijn, ja dan heb je wel een probleem. Want dan ben je geneigd oké, alles is groen, oké door. Ja dat is een goeie vraag, maar ja dan denk ik ook je moet ook altijd weer terug naar die patiënt van hoe ligt die erbij en gaat het goed?” (R3).

Op die manier kan dat wel eventueel gevaarlijk zijn als er daardoor dingen over het hoofd worden gezien, in het geval een arts de app zodanig vertrouwt en niet kijkt naar de waardes die bij de groen gekleurde organen horen. Zoals eerder aangegeven door een respondent zou de app er juist voor moeten zorgen dat er minder dingen over het hoofd worden gezien. De kritische blik van de arts blijft dus noodzakelijk.

Er wordt gesteld dat het vertrouwen van de arts in de app vanzelf zal komen wanneer de app zich verder ontwikkelt en er zoveel tests zijn gedaan dat het systeem accuraat genoeg is om op te vertrouwen (R4, R5, R6). Zij stellen dat er in samenspraak met artsen veel gevalideerd moet worden en bevestigd moet worden wanneer zij iets betrouwbaar genoeg vinden en waar de grens ligt. Respondent 6 benoemt dit als dat *“de app hem moet overtuigen daarvan”*. Daarvoor is heldere communicatie vanuit de app noodzakelijk. Het vertrouwen in de app is een proces van de arts met de app: *“En waarschijnlijk als de app het dan fout heeft, dan is hij de eerstvolgende tien keer gebruikt ‘ie hem niet. Of gebruikt ‘ie hem zijdelings langs de zijlijn en dan bouw je weer dat vertrouwen op dat de app het de volgende tien keer wel goed heeft” (R5)*. Daarin zal het vertrouwen afhankelijk zijn van hoe goed het model is, wat aansluit bij wat respondent 4 aangaf over de accuraatheid van het systeem. Daarbij wordt aangegeven dat het systeem voldoende vertrouwen moet geven om het te kunnen gebruiken, maar dat het niet zover komt dat je ‘er blind op kunt varen’ (R7). Er moet dus een middenweg worden gevonden tussen genoeg vertrouwen dat de app gebruikt wordt, maar ook genoeg ‘wantrouwen’ dat de arts kritisch blijft denken.

4.3 Interpretatie van data

De interpretatie van data door de ontwikkelaars is te duiden op verschillende punten in het ontwikkelingsproces, zoals het weergeven van de kleuren, het maken van de prognose, bias door missende data, ambiguïteit en onzekerheid van data.

Bijna alle respondenten geven aan dat de kleuren aangeven hoever iemand afwijkt van vergelijkbare historische patiënten, waarbij rood aangeeft dat iemand ver van vergelijkbare patiënten af ligt, maar niet per se in levensgevaar hoeft te zijn. Het wil meer zeggen dat er aandacht nodig is. Ook maakt de app een vergelijking in het beloop van de gezondheid van een patiënt met andere vergelijkbare patiënten. Dat de app dit vergelijking maakt, wordt als optimaler gezien dan dat de arts dit zelf doet, omdat er op een objectievere manier veel meer patiënten kunnen worden vergeleken (R1). Het beïnvloedt daarmee de interpretatie van de data door de arts doordat het al een suggestie doet van welke data bij elkaar hoort. Het gebied waar focust op moet zijn, wordt daardoor kleiner gemaakt (R1). Wat deze vergelijking lastig maakt, is tijd (R7). De patiënten liggen namelijk niet even lang in het ziekenhuis, dus *“wanneer weet je of een patiënt van dag 27 op een patiënt van dag 28 lijkt, dus waar ga je die afkapwaardes leggen? Welke patiënten laat je wel en niet op elkaar lijken?” (R7)*. Dit vormt nog wel een probleem. Ook het bepalen van de kleuren vormt op dit moment nog een probleem. De kleuren zeggen iets over de kwaliteit

van data, maar kunnen niet per casus worden afgewogen. *“Wat onze applicatie doet is dat het je ook nog een overweging meegeeft van wat normaliter in een bepaalde situatie wordt meegewogen”* (R7). Het is nog wel de vraag hoe dit moet worden ontwikkeld. Het vergelijken van de patiënten en het bepalen van de kleuren zijn fundamentele kwesties voor het ontwikkelen van de app en voor het creëren van (toegevoegde) waarde. Dat het nog onduidelijk is hoe dit precies moet worden gedaan en hoe er met de problemen moet worden omgegaan, maakt dat het waarde-creatie proces van de ontwikkelaars nog niet uitgekristalliseerd is.

Ook wordt door bijna iedereen aangegeven dat de dataset (voorlopig) enkel gebaseerd is op historische patiëntgevallen. De interpretatie van deze data hoeft en kan niet 100% goed zijn, het algoritme kan er af en toe naast zitten (R1). Het algoritme zal wel steeds accurater worden, omdat het zich baseert op steeds meer patiënten (R4). Toch zullen er altijd gevallen blijven waarin bijvoorbeeld een orgaan rood kleurt maar eigenlijk oranje had moeten zijn, dat is lastig om te voorkomen (R4). Met deze verwachting dat de kleuren wellicht niet altijd nauwkeurig zullen zijn, moet daarbij ook worden ingespeeld op hoe artsen daarmee omgaan en hoe dat wordt gecommuniceerd. Een van de problemen daarbij kan de ambiguïteit van medische data zijn en daarmee de discrepantie tussen de rationele logica van het algoritme en de subjectieve interpretatie die medische data nodig heeft om te kunnen duiden. De kleuren zijn een interpretatie van de data, dus is wel de vraag of die kleuren accuraat genoeg kunnen worden door middel van rationele logica om een goed beeld te kunnen geven.

De kleuren zijn een categorisering die aan te passen valt, zo stellen de meeste respondenten. Daarin is het de ambitie om die kleuren dynamischer te maken, *“[...] zodat een arts zelf kan aangeven binnen dit soort ranges zou ik dit groen willen hebben”* (R1). Met uiteindelijk dat dit ook op patiëntniveau kan. Zodat de arts zelf meer de interpretatie van de kleuren kan maken dan de computer. De kleuren hebben ook geen harde grens: *“het is vaak een combinatie van parameters en vaak is het ook nog per patiënt dat het verschilt”* (R3). Historische patiëntgevallen zijn een betrouwbare voorspeller, maar het is nooit genoeg om een perfecte voorspelling te doen (R6). Er zal altijd een arts nodig zijn die door studie of ervaring input kan geven. Op deze manier wordt het probleem dat in de vorige alinea geschetst is, gedeeltelijk opgelost. Een arts kan zodoende zelf aangeven wanneer kleuren niet kloppen en er zijn eigen subjectieve interpretatie aan geven. De mate waarin dit moet gebeuren is wel de vraag, want hoe zit het dan met het doel van de app? De kleuren moeten een indicatie geven van hoe het met de patiënt gaat, maar als deze kleuren dynamisch zijn, wat zegt het dan? Er wordt dan ook gesteld dat het nog wel lastig is om de betekenis van de kleuren te duiden (R3). Hoeveel waarde kan de arts eraan hechten of hoeveel zegt een getal? Daar komt bij dat er heel veel scenario's zijn waarin de data niet eenduidig is, omdat bijvoorbeeld een meting niet goed is gebeurd, een slangetje is afgekoppeld of er sprake is van oude data. Al die scenario's maakt het complex om een soort eindigheid aan de opties te geven. Dat soort dingen zijn lastig om te verwerken in de app. In gesprek met de arts zal dit betekenis moeten krijgen. Dit is een interessante bevinding, omdat met de kleuren de ontwikkelaars ook betekenis geven aan de data. Op welke manier moet de arts er dan ook betekenis aan geven en op welk moment in het proces? Aan de ene kant zou het het ambiguïteitsprobleem kunnen oplossen, aan de andere kant lijkt het ook dat de kleuren wellicht niet bepaald kunnen worden door de ontwikkelaars, omdat er zich te veel problemen bij voordoen die een rationeel systeem niet op kan lossen.

Sommige respondenten denken hier anders over. Er wordt gesteld dat de interpretatie van het algoritme uiteindelijk accurater zal zijn dan de interpretatie die een arts zal maken, doordat het algoritme getraind is op duizenden patiënten (R4). Daarnaast kan de bias van een algoritme gemakkelijker aangepast worden dan die van een mens:

“Ik denk dat uiteindelijk de gedachte is hetzelfde. Je gebruikt gewoon je ervaring, dat is in de app je model die op basis van historische data een beslissing maakt. En bij de arts is het zijn ervaring überhaupt met historische patiënten. Dus die gedachtegang is een beetje hetzelfde. [...] Alleen is een model, is een computer dus die denkt niet na over wat mijn twee vorige patiënten waren, maar die denkt over de vorige 500 patiënten waardoor je minder onzekerheid krijgt en minder bevooroordeeld of bias door zeg maar recente patiënten” (R5).

De app is in die zin ‘objectiever’ dan een arts. Dit wordt verduidelijkt met een voorbeeld van een arts die vertelde dat hij bij een kind een bacterie gemist had (R5). Daardoor gaf hij bij de volgende tien kinderen sowieso antibiotica als hij maar enigszins vermoedde dat het kind ziek was. Terwijl een algoritmisch model dit niet meeneemt. Wat wel een aandachtspunt is, is dat er vooraf bepaalde parameters worden bepaald die als belangrijk worden geacht, terwijl een arts een parameter kan zien die veel effect kan hebben, maar niet in het model wordt meegenomen (R6).

Er worden in principe dus tegenstrijdige uitspraken gedaan. Aan de ene kant zou de app accurater zijn dan een arts, maar aan de andere kant zou de app wel dynamisch moeten worden om de interpretatie van de arts mee te nemen. Hieruit kan opgemaakt worden dat er nog geen duidelijkheid is over hoeveel waarde de ontwikkelaars toevoegen aan de data en hoeveel ze kunnen toevoegen. De ambitie is dat de app accurater zal zijn dan een arts, maar of dit realistisch is, is de vraag. Dit zou wel overeenkomen met wat uit het theoretisch kader blijkt. Door de ambiguïteit van medische data en de noodzaak van de subjectiviteit (ervaring, intuïtie) van artsen, is het lastig te bepalen waar waarde kan worden toegevoegd met rationele logica.

4.3.1 Bias door missende data of artefacten

Bias door missende data of artefacten betekent dat er bijvoorbeeld metingen missen of fouten zitten in de metingen door bewegingen van patiënten, apparatuur die niet goed is aangesloten, operaties die plaatsvinden, et cetera. Voor de ontwikkelaars is het lastig om hier mee om te gaan, omdat ze er soms niet achter komen dat dit aan de hand is of er niets mee kunnen (R1, R2). Om dit op te lossen wordt geopperd om extra variabelen mee te kunnen nemen of het journaal (geschreven tekst over de patiënt door artsen) te kunnen gebruiken. Op die manier is het mogelijk om in de tijd te zien wat er met de patiënt gebeurt, wanneer er bijvoorbeeld operaties plaatsvinden en de metingen dus niet kloppen. Het probleem daarbij is dat voor het modelleren het lastig is om zoiets handmatig te doen voor duizenden patiënten. Daarnaast is het ook lastig om waardes te herkennen die niet kloppen, terwijl ze er normaal uit zien:

“Een hele rare waarde, die zijn er wel makkelijk uit te halen maar het kan dus ook zijn dat je bijvoorbeeld een normale waarde ziet maar die dan gewoon 20 minuten lang, bij wijze van spreken, constant blijft. Waarvan je denkt oké zou kunnen, maar...” (R2). “Dit is wel heel erg raar” (R1). “Dat je exact hetzelfde blijft en dan blijkt dat op dat moment er iets in de infuuszak, of dat een infuus gespoeld moet worden waardoor de meting dus niet goed verloopt” (R2).

Wanneer er niet genoeg of geen data beschikbaar is, dan zal de app dat ook communiceren door te zeggen dat er geen data beschikbaar is (R4). Hoewel dat voor de hand liggend lijkt, blijkt ook uit bovenstaand stuk dat het niet altijd te herkennen is. In die gevallen zal de arts dit moeten opmerken, maar met de suggestie die de app al geeft over hoe het met de patiënt gaat kan dit ook lastig te herleiden zijn. Een ander lastig punt is dat er veel verschillende databronnen zijn en niet alle data vindbaar is, doordat bijvoorbeeld artsen en verpleegkundigen iets doen maar het niet opschrijven (R5, R6). Daarnaast heeft ook iedereen een eigen manier van werken, waardoor soms ook informatie niet te vinden is. In de app wordt geprobeerd om wel één systeem te creëren waar alles samenkomt. Ten slotte is er ook het probleem van de beslissingen die artsen al maken voor de ruwe data. Zij besluiten namelijk om bijvoorbeeld een patiënt aan te sluiten op een hartslagmonitor (R1). Daar zit al een keuze achter van de

artsen om die data wel of niet op te vragen. Op die manier is de beschikbare data ook al gekleurd, omdat daar al keuzes achter zitten.

Missende data of artefacten kunnen dus een grote rol spelen in de waarde die de ontwikkelaars toevoegen in de app. Hoe wordt dit ondervangen en in het geval dat dit niet lukt, hoe wordt dit dan gecommuniceerd? Hoe blijft de arts kritisch genoeg om alert te zijn op deze problemen, maar behoudt hij of zij het vertrouwen in de app?

4.3.2 Ambigüiteit van data

De ontwikkelaars zijn nog niet bezig met het vraagstuk ambigüiteit van data. Het model laat zien hoever een patiënt van vergelijkbare patiënten aflight en hoe het verloop van zijn of haar gezondheid zou kunnen gaan (R1, R2). Uiteindelijk kunnen er gewichten worden gehangen aan welke parameters of variabelen belangrijk zijn. Er wordt gesteld dat dit volledig vanuit de artsen zou moeten komen in het ontwikkelproces (R1, R2). Daarin zou bijvoorbeeld een gradatiesysteem gemaakt kunnen worden waarin de belangrijke variabelen worden onderscheiden van minder belangrijke variabelen, waarbij de belangrijke variabelen meer meetellen dan de minder belangrijke variabelen in het bepalen van de gezondheid. Daarnaast is de context ook erg belangrijk: bepaalde waardes kunnen voor een kind van 2 jaar heel anders geïdentificeerd worden dan voor een kind van 2 weken oud. Ook wordt benoemd dat missende data een vorm van ambigüiteit is (R7). Hoe de waarde van data geïdentificeerd moet worden zal dus in samenspraak zijn met artsen.

Er wordt wel gezien dat ziektebeelden of andere externe invloeden van invloed zijn op hoe de data eruitziet. Saturatie-waardes voor kinderen met hartproblemen zien er anders uit dan voor kinderen met een normale bloedsomloop. Zulke ambigüiteit is voor een computersysteem lastig in te schatten waardoor artsen wel nodig zijn:

“Ja ik denk dat we dan uiteindelijk toch weer terugkomen bij het gedeelte dat de artsen zelf nog een bepaling kunnen maken waar ze bepaalde cut-offs zouden willen zien. Want het computersysteem zelf gaat niet zo heel snel ambigüiteit identificeren en ambigüiteit inschatten. Dat is uiteindelijk toch iets waar een mens beter in is dan een machine die alleen maar in wel of niet denkt. Dus ik denk dat vooral in dat gedeelte het stukje ambigüiteit terecht gaat komen” (R1).

Daarbij zouden wel default instellingen gemaakt kunnen worden, die de arts vervolgens zelf kan aanpassen. Daarbij wordt benoemd dat menselijke interpretatie onmisbaar blijft:

“Dus er is eigenlijk externe informatie, dus ambigüiteit kun je alleen maar oplossen met begripsmatige principes. Dus je kunt eigenlijk alleen maar, tenminste dat is de enige oplossing, dus door menselijk redeneren. [...] Tweede is dat er ook een bepaald begrip van de hele situatie in zit die niet in die applicatie staat, deze applicatie beoogt juist niet per se om die ambigüiteit volledig in beeld te brengen, dat zou niet gaan want dan zou het eerder onzekerheid zijn, het gaat er juist over dat de enige persoon die eigenlijk iets met die ambigüiteit kan is die mens. En die kan die misschien toevoegen aan die app door bijvoorbeeld te zeggen wat normaal een referentie iets is” (R7).

Hoewel het lijkt daarmee een oplossing te zijn gevonden voor het ambigüiteitsprobleem, zou er ook gesteld kunnen worden dat wat de arts moet toevoegen aan referentie, het hele idee is van het kleurensysteem.

Wat daarnaast door elke respondent genoemd wordt, is specifiek de moeilijkheid van het bepalen van de kleuren, waar ambigüiteit een grote rol speelt. Hoe de kleuren bepaald moeten worden is geen harde, universele grens voor te stellen want heel veel parameters, zoals ziektebeeld en leeftijd van de patiënt,

bepalen wat de kleur moet zijn. Een van de respondenten geeft aan er vertrouwen in te hebben dat wanneer het algoritme maar op genoeg patiënten getraind is, het ook steeds accurater zal worden. Maar er zullen altijd speciale gevallen blijven waarin een orgaan bijvoorbeeld rood kleurt maar oranje had moeten zijn (R4). De oplossing daarvoor volgens deze respondent is dat de arts en het algoritme samenwerken. Hoe dit precies zou moeten gebeuren, wordt niet aangegeven. Dit komt overeen met wat eerder werd gezegd. Aan de ene kant lijken de respondenten van mening te zijn dat het systeem accurater zal zijn dan een arts, maar dat tegelijkertijd de interpretatie van de arts ook nodig is om het allemaal te kunnen duiden. De moeilijkheid van het duiden van de data blijkt ook uit een voorbeeld dat wordt gegeven: als het met een patiënt stabiel slecht gaat, krijgt het dan het label ‘goed’ (R5)? Hoe er uiteindelijk met ambiguïteit om moet worden gegaan, is niet duidelijk. De ene respondent stelt dat het model om moet gaan met de ambiguïteit (R5), de ander stelt dat het in samenspraak moet gaan met de arts (R4). Duidelijk is in ieder geval dat er dus nog weinig duidelijkheid over is. Wat uit het theoretisch kader bleek dat ambiguïteit een probleem kan vormen dat een algoritme of een rationeel systeem niet kan oplossen, blijkt wel deels uit de interviews. Er is verdeeldheid over hoe het aangepakt moet worden en de input van de arts wordt als belangrijk geacht. Hoe het uiteindelijk uitgewerkt moet worden, is niet duidelijk.

4.3.3 Onzekerheid van data

De respondenten geven aan dat de voorspelling van de gezondheid van de patiënt vooral gaat over het laten zien van hoe het verloop van de gezondheid van vergelijkbare historische patiënten was, om op die manier inzicht te krijgen in hoe het zou kunnen verlopen met de huidige patiënt. Hoe meer vergelijkbare patiënten er zijn, hoe accurater de voorspelling zal zijn (R1, R2). Daarbij is het belangrijk dat er niet te ver in de toekomst wordt gekeken, en dat er goed gecommuniceerd moet worden hoe zo een voorspelling gelezen kan worden (R3). Er wordt aangegeven dat het verschil met nu is, dat een arts ook historische patiënten uit zijn ervaring kent, maar de app zich kan baseren op veel meer historische patiënten waardoor er minder onzekerheid en bevooroordeeldheid in de voorspelling zit (R5). Ook wordt aangegeven dat de voorspelling van de gezondheid van de patiënt ook iets kan zeggen over de actuele status van de patiënt, en hoe zekerder die voorspelling is, hoe beter beeld de arts van de gezondheid van de patiënt krijgt (R7). De onzekerheid wordt groter naarmate er informatie ontbreekt. Met onzekerheid kan dus beter worden omgegaan door de ontwikkelaars dan met ambiguïteit, omdat dit goed kan worden weergegeven en gedetecteerd in de app. Dit was ook de verwachting uit het theoretisch kader.

4.4 Samenvattend

Met de interviews is getracht een antwoord te vinden op de deelvraag *Hoe wordt waarde toegekend aan medische data door de ontwikkelaars van UrSmartStatus?* Uit de gesprekken en aan de hand van de vragenlijst kwamen drie grote thema's naar voren: het verschil tussen de app en het EPD, besluitvorming en interpretatie van data.

Vershil app en EPD

Over het verschil tussen de app en het EPD kan gezegd worden dat het werkproces van de arts meer centraal komt te staan bij het gebruik van de applicatie. Dat betekent dat er aandacht is voor hoe mensen denken en hierop wordt ingespeeld middels de presentatie van de data. De indeling per orgaan is volgens de ontwikkelaars een logischere indeling voor de artsen dan de tabblad-indeling van het EPD waarin het meer ‘input’ gericht is zoals een tabblad lab-uitslagen. Daarnaast worden er conclusies verbonden aan de data aan de hand van het kleuren van de organen, gebaseerd op de vergelijking met historische patiënten. Daarvoor wordt big data gebruikt, waar in het EPD geen sprake van is. Op die manier wordt er al een deel van de interpretatie gemaakt door de ontwikkelaars/het algoritme. De data wordt in de app visueel samengevat en krijgt prioritering. Dit wordt ook wel gezien als een top-down manier van denken,

waarin eerst een globaal beeld wordt gegeven van de patiënt en de arts vervolgens gedetailleerder naar de specifieke waardes kan kijken. Anders dan in het EPD, waarin het meer bottom-up gericht is en er vanuit de specifieke waardes een beeld moet worden gevormd die vaak al vanuit een bepaalde verwachting wordt gecreëerd.

Volgens sommige van de ontwikkelaars hoeft deze app niet veel anders te zijn dan andere instrumenten die ingezet worden. Er zit volgens hen alleen meer data en technologie achter. De ontwikkelaars zijn van mening dat er genoeg ruimte voor interpretatie overblijft voor de arts bij de app. Hoe dit tot uiting zal komen is niet duidelijk. De arts wordt met de prioritering in een bepaalde richting gestuurd, maar heeft volgens de ontwikkelaars wel ‘genoeg’ ruimte voor eigen interpretatie. Hoe verhoudt zich dit tot elkaar en wat is genoeg? Daarnaast werd in het theoretisch kader duidelijk dat het inzetten van algoritmes implicaties kan hebben voor hoe de data geïnterpreteerd wordt (Mittelstadt et al., 2016). Met de verwerking van data middels algoritmes worden er keuzes gemaakt en waardes gereflecteerd waardoor de artsen in principe een subjectieve selectie krijgen van de data. Op welke manier wordt hierover gecommuniceerd en in hoeverre intervenueert dit met de klinische besluitvorming van de arts?

Besluitvorming

Zoals in het theoretisch kader aangegeven, zou de besluitvorming van de artsen kunnen veranderen door het gebruik van een decision-support systeem, omdat daarin enkel een rationele logica wordt gevolgd en een arts gebruikmaakt van zowel de rationele als klinische logica. De ontwikkelaars zien deze rationaliseringsslag als een verbetering van de rationele logica van de artsen: een computer kan een betere, objectievere vergelijking maken op basis van veel meer historische patiënten dan dat een arts dat kan doen. Mensen hebben namelijk een beperkte capaciteit in het geheugen en zijn sensitief voor de recentheid van herinneringen en herinneringen die passen bij de eigen overtuigingen (March, 2010, p.45). Het algoritme mist echter de intuïtie en klinische ervaringskennis die een arts wel heeft, maar doet in principe wel een suggestie voor de richting waarin gedacht moet worden. Kan die richting al worden gegeven op het moment dat dit enkel gebaseerd is op rationele logica? Er wordt gesteld dat de applicatie ook dynamisch zou moeten zijn waardoor de arts interactie heeft met de app. De arts zou de kleuren kunnen aanpassen wanneer hij of zij dit nodig acht. De ontwikkelaars geven dan ook aan de interpretatie van de arts erg belangrijk te vinden. Hier vindt een discrepantie plaats: de arts moet vertrouwen hebben in de applicatie om er mee te kunnen werken en voor een eerste indruk af kunnen gaan op de kleuren, maar blijft wel eindverantwoordelijke en is de enige in het proces die intuïtieve en ervaringskennis toevoegt. Op welke manier de arts precies zou moeten omgaan met de app is dus nog niet volledig duidelijk. Zo wordt ook gesteld dat de kleuren moeten ‘uitnodigen om verder te zoeken’, maar hoe wordt dit gecommuniceerd naar de arts? En hoe verhoudt zich dat tot het geven van een snelle, eerste indruk van de gezondheid van de patiënt? En hoe zit het met gevallen wanneer het orgaan groen kleurt, maar eigenlijk rood had moeten zijn?

Interpretatie van data

Wat betreft de interpretatie van de data wordt gesteld dat er uiteindelijk de ambitie is dat de applicatie accurater zal zijn dan een arts. Dit zou moeten voortkomen uit het vergelijk dat het algoritme kan maken met enorme hoeveelheden aan historische patiëntgevallen, terwijl een arts daar het geheugen niet voor heeft en bevooroordeeld is in zijn of haar beoordeling daarvan. Toch zou de app wel genoeg ruimte moeten bieden voor de interpretatie van de arts door het dynamisch te maken. Het lijkt erop dat er getracht wordt het rationele gedeelte van besluitvorming beter te maken dan de mens, en de mens te laten doen waar hij goed in is. Gezien de ambiguïteit van medische data is het wel de vraag of dit mogelijk is, aangezien rationele en klinische besluitvorming niet per se iets is wat op elkaar volgt maar met elkaar verweven is. De vergelijking met historische patiënten zou gezien kunnen worden als iets

wat ook met de ervaring en intuïtie van de arts gedaan moet worden, omdat die de ambiguïteit van informatie kunnen interpreteren. Missende data of artefacten kunnen ook gezien worden als een vorm van ambiguïteit. Het is nog onduidelijk of dit kan worden ondervangen door de app en zo niet is de vraag wat de waarde kan zijn van de kleuren. In dat geval moet een arts altijd kritisch blijven op wat de app suggereert, maar kan er dan nog wel vertrouwen zijn in het systeem? De input van de arts zal in ieder geval belangrijk blijven in zowel het ontwikkelproces als bij het gebruik, zo wordt ook door de ontwikkelaars gesteld. Met onzekerheid van data kan door de ontwikkelaars wel goed worden omgegaan, omdat een gebrek aan informatie die duidelijk is gecommuniceerd kan worden naar de arts. Zodoende kan de onzekerheid van een voorspelling of status in beeld worden gebracht. Dit komt overeen met de verwachting uit het theoretisch kader, omdat onzekerheid iets is wat in getallen uit te drukken is.

5. Resultaten artsen

Om de empirische deelvraag *Hoe wordt waarde toegekend aan informatie uit de app UrSmartStatus door medici?* te beantwoorden, is er een experiment gedaan met zes artsen waarbij op voorhand twee artsen het experiment testten om eventuele fouten te kunnen corrigeren. Dit bleek te gaan om kleine aanpassingen waardoor de bevindingen van de artsen in de testfase zijn meegenomen in het resultaat. Om antwoord te kunnen geven op de deelvraag, is er in het experiment gekeken naar de verschillen tussen de interpretatie van de arts met gebruik van de app en met gebruik van het EPD. Daarnaast is er gevarieerd in mate van ambiguïteit aan de hand van twee verschillende patiëntcasussen. Om deze verschillen te kunnen duiden zijn er concrete hypothesen opgesteld waarin de verschillen kunnen worden getoetst:

H1: De medicus kan met minder moeite een interpretatie geven aan de data met het gebruik van de app dan met het gebruik van het EPD.

H2: De medicus komt tot een uitgebreidere differentiaal diagnose met gebruik van de app dan met gebruik van het EPD.

H3: Hoe hoger de mate van ambiguïteit van de casussen, hoe meer voordeel de app geeft voor de interpretatie van de data door de medicus ten opzichte van het EPD.

Daarnaast is er gelet op punten uit het theoretisch kader waarop verwachtingen waren opgesteld en andere opvallendheden uit de gedachteprocessen en follow-upvragen.

5.1 Beschrijving experiment en analyse

Het experiment verliep als volgt. Ieder van de artsen kreeg een uitleg over het experiment (bijlage D) waarna ze hun gedachteproces bij het stellen van een diagnose en beleid uitspraken bij twee casussen. Deze casussen kwamen de ene keer in de app, de andere keer in het EPD waarbij ook de volgorde van de casussen gewisseld werden. Het hardop uitspreken van het gedachteproces werd samen met de acties op het scherm opgenomen. Ter illustratie een fragment van een hardop denkproces:

“Nou dan klik ik op het longsymbooltje. Dan zie ik een saturatie, even kijken hoor, bovenaan een aantal tijden. Saturatie, ademhalingsfrequentie, ademarheid, intrekken luchtweg, en een bloedgas en een X-Thorax oke nouja hij heeft een saturatie die een beetje tegenvalt, maar met een DORV inlet VSD zou ik niet meteen alarm alert op 89% met wat mengcirculatie waarschijnlijk. Heeft ie een ademfrequentie van 65 per minuut, dat is wel aan de ruime kant bij een kind van deze leeftijd. Ik ga zo nog even naar de trend kijken. Heeft een toegenomen ademarheid met intrekking subcostaal en is die snotterig en hoesten. Nou dat zou kunnen passen bij een infectieus verhaal maar ook met dit verhaal bij overvulling met z'n cardiale beeld”
(Arts1)

Bij het gebruik van de app kon de arts ook in het EPD kijken wanneer hij of zij dat wilde. Dat EPD was hetzelfde ingericht als de versie van het EPD zonder de app, dus er kon ook informatie worden gevonden die ook in de app te vinden was. Daarnaast stonden er tabbladen in met informatie die niet in de app waren te vinden, zoals een X-Thorax foto.

De transcripten zijn middels een script analyse beoordeeld. De volgende operators konden worden onderscheiden:

<i>Operator</i>	<i>Betekenis</i>
<i>Bestuderen</i>	Data of informatie lezen
<i>Besluiten</i>	Kiezen om een actie te verrichten (bijv. “ik kijk nog even naar de longen”)
<i>Vaststellen</i>	Waarde geven aan iets maar niet aan medische data (bijv. “recht ingeschoten foto” of “hier kom ik niet verder mee”)
<i>Afvragen</i>	Zichzelf te rade gaan
<i>Verklaren</i>	Een reden geven
<i>Vragen</i>	Aan interviewer een vraag stellen
<i>Diagnosticeren</i>	Tot een (gedeeltelijke) diagnose komen
<i>Beoordelen</i>	Oordeel geven aan medische data/waarde (bijv. “7.2 dat is goed”)
<i>Samenvatten</i>	Al genoemde informatie nog een keer compact noemen
<i>Beleid</i>	Noemen van beleid (bijv. “ik zou Optiflow toedienen”)

Verder is vastgelegd hoe lang een arts deed over het gedachteproces bij een casus, op welke momenten geswitcht werd tussen de app en het EPD en wat hun differentiaal diagnose was.

Het experiment is zo opgesteld dat het voor veel externe invloeden heeft kunnen corrigeren. De artsen kregen altijd de twee verschillende casussen aangeboden, de ene keer in de app en de andere keer in het EPD, waarbij de volgorde van aanbieden ook varieerde. Deze variatie werd vooraf aangebracht zonder dat bekend was welke arts welke variant zou krijgen. De verdeling van app en EPD en de volgorde over de artsen kan dus als ‘at random’ worden beschouwd. Zodoende kon de kans worden geminimaliseerd dat bij het vergelijken de resultaten een gevolg zijn van specifieke personen, de casussen, de volgorde of tijdsduur.

5.2 Verschil in hoeveelheid geobserveerde informatie tussen app en EPD

In deze analyse is gekeken naar de hoeveelheid informatie die de artsen observeerden aan de hand van de app en aan de hand van het EPD. Daarvoor werd een lijst met items opgesteld waarbij elk item een stukje informatie representeerde, zoals het bekijken van de ademprequentie of de vochtbalans. Daarbij is gekeken naar de verschillen per casus tussen het gebruik van de app en het EPD.

<i>Casus</i>	<i>App of EPD</i>	<i>Aantal items</i>	<i>Aantal geobserveerde items</i>	<i>Kans op geobserveerd</i>
1	App	40	36	0.9
1	EPD	40	25	0.625
2	App	30	21	0.7
2	EPD	30	25	0.833

Tabel 1 laat de resultaten zien van de vergelijking per casus per getest systeem.

In casus 1 werden meer items geobserveerd bij gebruik van de app dan bij gebruik van het EPD. Wanneer de artsen de app gebruikten, vergaarden ze dus meer informatie om tot hun diagnose te komen dan wanneer ze het EPD gebruikten. Bij casus 2 is dit andersom: daarin werden minder items geobserveerd bij gebruik van de app dan bij gebruik van het EPD, wat betekent dat ze minder informatie uit het de app haalden dan uit het EPD om tot hun diagnose te komen, al is het verschil kleiner dan bij casus 1.

5.3 Verschillen in besluitvormingsproces

Door te kijken naar de verschillen in het gedachteproces bij het gebruik van de applicatie en bij het gebruik van het EPD kan er iets gezegd worden over de verandering in het interpretatieproces. Dit kan deels geduid worden door te kijken naar de hoeveelheid moeite die de artsen moeten doen om een interpretatie te geven aan de data. De volgende hypothese is daarvoor opgesteld:

H1: De medicus kan met minder moeite een interpretatie geven aan de data met het gebruik van de app dan met het gebruik van het EPD.

Er is gekeken naar momenten waarop de artsen tot een (gedeeltelijke) diagnose kwamen (is dat vroeg of laat in het proces?), momenten waarop de artsen die de app gebruikten het EPD zijn gaan gebruiken (missen ze informatie in de app of is het EPD gemakkelijker?), de snelheid van het gedachteproces en hoeveel ze bestudeerden voordat ze tot hun uiteindelijke diagnose kwamen. Daarnaast is bekeken op welke manier hun antwoorden op de follow-upvragen verklaringen kunnen voor de moeite van hun interpretatieproces en tevens of die antwoorden overeenkomen met hun gedrag tijdens het gedachteproces.

Momenten van switchen app naar EPD en momenten van diagnose stellen

Door te bekijken op welke momenten een arts switcht van de app naar het EPD en te kijken naar de momenten waarop de arts tot een (gedeeltelijke) diagnose komt, kan dit iets zeggen over in hoeverre ze met behulp van de app tot een diagnose kunnen komen en in hoeverre ze het EPD nog nodig hebben voor hun interpretatie.

In onderstaande twee tabellen zijn de momenten te zien waarop artsen die een casus beoordelen in de app switchen naar het EPD om daar informatie te bekijken. In de tabel is weergegeven om welke arts het gaat, om welk moment het gaat en of het nodig was of de informatie ook in de app te vinden was. Met nodig wordt bedoeld dat de informatie niet in de app te vinden is, zoals de X-Thorax foto. Die staat enkel in het EPD. Met ‘app’ wordt bedoeld dat de informatie zowel in de app als in het EPD te vinden is. Dat betekent dat ze informatie hebben opgezocht in het EPD, maar dit ook te vinden was in de app. Daarnaast is bij opmerkingen informatie te vinden aangaande de interpretatie en diagnose van de arts en wanneer het voorkwam ook opmerkingen over de trend, waar later aandacht aan zal worden besteed.

Casus 1

Arts	Switch	Moment	Noodzaak	Opmerkingen
Arts 1	1 ^e	Radiologie-X-Thorax: arts had gezien in de app dat deze verricht was	Nodig	Voor diagnose niet behulpzaam
	1 ^e	Lab: om trend te zien van bloedgas	App	Wou terug naar app maar ziet lab staan en kijkt daar eerst
				Noemt het klikken op het tijdspad in de app ‘de trend’
				Komt al tot diagnose voor het gebruiken van EPD, trekt alleen conclusie uit EPD dat het geen respiratoire acidose is

Arts 2	1 ^e	Radiologie-X-Thorax: arts had gezien in de app dat deze was verricht	Nodig	Trekt er geen conclusie uit
	1 ^e	Voorblad: journaal bekijken om beoordeling te bekijken. Leest dat vooral ademhaling probleem is en past bij Rhino virus	App	Maakt hier gedeeltelijke diagnose (virusinfectie)
	1 ^e	Lab: temperatuur, CRP, BNP en bloedgas bekijken	App	Sluit bacteriële infectie uit door temperatuur
	1 ^e	Respiratie: noemt niets van het tabblad	App	
	1 ^e	Circulatie: noemt niets van het tabblad	App	
	1 ^e	Virologie: ziet Rhino virus positief	App	Gedeeltelijke diagnose (virusinfectie)
	1 ^e	Circulatie: saturaties bekijken	App	
				Komt pas tot diagnose in EPD (maar veel van de waardes ook te vinden in app). Vraagt in app wel af of ze niet overvuld is.
Arts 6	1 ^e	Radiologie-X-Thorax: arts had gezien in de app dat deze was verricht. Denkt aan infectie dus wil lab zien	Nodig	Gedeeltelijke diagnose (virusinfectie of atelectase)
	1 ^e	Lab: CRP bekijken, beeld van infectie dus wil virologie zien	App	Gedeeltelijke diagnose (virusinfectie)
	1 ^e	Virologie: Rhino virus positief, denkt aan bacterieel superinfect, wil terug naar app maar gebruikt toch Metavision	App	Gedeeltelijke diagnose (virusinfectie met eventueel bacterieel superinfect)
	2 ^e	Respiratie: intrekkingen, secreet en ademhalingsfrequentie bekijken	App	Bevestiging van diagnose
	2 ^e	Circulatie: bevestigt waar arts aan dacht	App	Bevestiging van diagnose
	2 ^e	Lab: checken of dat ook van de 18 ^e was, CRP bekijken of het van nu is	App	
	3 ^e	Medicatielijst: bij noemen van DD met circulatoir probleem medicatielijst checken, kon niet direct vinden in app	App	Gedeeltelijke diagnose (hartfalen)
				Komt tot diagnose in EPD (maar wel veel waardes ook te vinden in de app)
Arts T1	1 ^e	Radiologie-X-Thorax: arts had gezien in de app dat deze was verricht	Nodig	Ziet wellicht beetje overvulling
	2 ^e	Virologie: bekijken wanneer dat voor het laatst is gedaan (had al diagnose gegeven van virale infectie)	App	

	2 ^e	Lab: had het eerder allemaal al gezien, merkt op dat BNP eerder veel hoger was	Nodig (zo ver terug gaat app niet)	BNP was ook al eerder opgemerkt in app
				Komt tot diagnose in app, ziet overvulling bij Thorax foto

Wat uit deze tabel blijkt is dat elke arts in eerste instantie naar het EPD switcht om te kijken naar de Thorax foto. Allen hadden al in de app gezien dat deze verricht was. Vervolgens wordt veelal in dezelfde switch gekeken naar andere tabbladen, waarvan de informatie ook in de app te vinden was. Dit kan verklaard worden doordat het gemakkelijk is om in hetzelfde systeem verder te klikken naar andere informatie en ze dat systeem ook gewend zijn. Daarnaast zou het ook zo kunnen zijn dat ze in hun hoofd een bepaalde volgorde van informatie zoeken hebben (omdat ze bijvoorbeeld een bepaalde diagnose in hun hoofd hebben en deze willen checken) en deze afgaan. Omdat ze het EPD gewend zijn, kan het zijn dat ze ook in dat systeem denken en dus niet snel teruggaan naar de app.

Ook bij de verdere switches is de informatie die wordt gezocht veelal te vinden in de app. Een verklaring kan zijn dat er op dat moment specifieke informatie wordt gezocht en een automatische reactie is om dat op te zoeken op de plek die ze kennen. In de gesprekken werd ook aangegeven dat de meeste artsen nog wel moeten wennen aan de app en dat ze niet precies weten welke informatie in de app is te vinden. Een andere verklaring, die met de eerste te maken heeft, kan zijn dat ze de app al gedeeltelijk zijn doorlopen en dan nog op zoek zijn naar andere informatie, en in de veronderstelling zijn dat deze niet in de app is te vinden. Nog een andere verklaring die ook met de gewenning aan de app te maken heeft, is dat de symbolen in de app wellicht niet voor zich spreken over de informatie die er te verkrijgen is.

Twee van de respondenten komen tot een diagnose in de app, de twee andere respondenten komen tot een diagnose in het EPD. Dat zou kunnen betekenen dat zowel de app als het EPD zo goed als los van elkaar gebruikt kunnen worden om tot een diagnose te komen (uitgezonderd uiteraard van informatie zoals foto's, die op dit moment in het EPD worden weergegeven maar in principe ook weergegeven kunnen worden in de app). Dat twee van de respondenten tot hun diagnose komen in het EPD, zou wellicht een gevolg kunnen zijn van het feit dat ze de X-Thorax foto bekijken en vervolgens doorklikken in het EPD uit automatisme.

Casus 2

Arts	Switch	Moment	Noodzaak	Opmerkingen
Arts 3	1 ^e	Vochtbalans: arts wil weten wat urineproductie en vochtbalans is	App	Vochtbalans niet gevonden in app
	1 ^e	Metavision overzicht: urineproductie per kilo per uur bekijken	App	
	1 ^e	Vochtbalans: bekijkt niets	-	
	1 ^e	Circulatie: effect bekijken op circulatie, ademfrequentie en bloeddrukken bekijken	App	
	2 ^e	Voorblad: bekijkt niets	-	

	3 ^e	Lab: arts wil weten wat er eerder aan lab is gedaan, na al diagnose te hebben gesteld	App	
				Komt tot diagnose in app, gebaseerd op natrium gevonden in app
Arts 4	1 ^e	Voorblad: zoekt operatieverslag, leest recentste anamnese en beleid	App	
	2 ^e	Vochtbalans: arts bekijkt wat er in gaat aan vocht	Nodig	Vochtbalans in app gezien maar wil weten wat er in is gegaan aan vocht
	2 ^e	Circulatie: bekijkt hartactie en bloeddruk	App	
	2 ^e	Anesthesieverslag 1 en 2: bekijkt welke vulling patiënt heeft gehad op de OK, bekijkt het aantal ringers. Wil het verslag lezen wat er gedaan is maar kan niet geopend worden	Nodig	
	2 ^e	Voorblad: wil bekijken hoe ze er circulatoir bijligt en wat haar capulair refill is. Bekijkt ook welke medicatie patiënt krijgt	App	Komt tot diagnose (ADH of te veel vocht)
	3 ^e	Lab: zou lab prikken om elektrolyten en urine te controleren dus bekijkt elektrolyten en nierfunctie, ook Hb	App	Duidelijkere diagnose (met name te veel vocht)
				Deels aanwijzingen uit app, deels aanwijzingen uit EPD voor diagnose maar stelt diagnose in EPD
Arts 5	1 ^e	Lab: wil trend in het lab bekijken	App	Nog een keer bekijken met trend. Dus wil toch trend zien
	1 ^e	Respiratie: bekijkt of ze echt geen ondersteuning heeft	App	
	1 ^e	Lab: bekijkt bloedgas, kalium en creat	App	Ook al gezien in de app
	1 ^e	Voorblad: leest journaal met informatie over intubatie, bekijkt medicatie		
	2 ^e	Lab: wil natrium urine weten, kan niet vinden	App	
	2 ^e	Voorblad: ziet alleen verslag van OK en heeft arts niets aan	Nodig	
	2 ^e	Metavision overzicht: zegt niets	-	
	2 ^e	Circulatie: noemt dat ze dat allemaal in de app zag	App	

	2 ^e	Vochtbalans: ziet dat die van positief naar negatief gaat	App	
	2 ^e	Klikt alle tabbladen aan maar zegt er niets over	-	
	3 ^e	Vochtbalans: zoekt wat er aan infusen loopt maar kan het niet vinden door gebrek aan ervaring met Metavision	Nodig	
				Diagnose deels in app en deels in EPD

Wat uit deze tabel blijkt, is dat het switchen naar het EPD voor deze artsen erg verschillend is van elkaar. Niemand begint met hetzelfde tabblad en de volgorde van informatie verkrijgen is ook verschillend. Wat wel overeenkomt is dat alle artsen drie keer switchen naar het EPD. Dit laatste zou kunnen betekenen dat ze veel informatie uit het EPD vonden nodig te hebben. Toch was veel informatie wel in de app te vinden. Ook hier gelden dezelfde verklaringen als bij casus 1: dat ze naar het EPD switchen terwijl de informatie ook in de app staat, kan gewenning van het systeem zijn, gebrek aan kennis over de informatie in de app, en het gemak van het kijken naar een ander tabblad wanneer ze al in het EPD kijken. Ook is te zien dat er met specifieke informatiebehoefte naar een tabblad in het EPD wordt geswitcht. Toch wordt er ook zonder doel naar de tabbladen gekeken; arts 5 klikt bijna alles aan maar haalt daar niet per se informatie uit.

Bij casus twee is het stellen van de diagnose wat onduidelijker dan in casus 1. Een van de artsen stelt de diagnose in de app, de twee anderen vormen hun diagnose deels uit de app en deels uit het EPD. De informatie die gebruikt wordt uit het EPD is echter ook allemaal te vinden in de app. Er wordt bij casus 2 wel meer geswitcht naar het EPD. Wat hier de oorzaak van zou kunnen zijn, is lastig te zeggen. Een aantal van de respondenten gaf aan minder ervaring te hebben met casus 2 dan met casus 1, wat de casus moeilijker kan maken. Twee van de drie artsen zoekt op het voorblad naar informatie over de patiënt en de operatie. Het kan zijn dat ze meer informatie nodig hebben, omdat ze niet bekend zijn met een dergelijke casus en daardoor meer informatie nodig hebben om tot een goed beeld te kunnen komen.

Snelheid

De snelheid waarmee de artsen de casus behandelden en tot een diagnose kwamen, kan ook iets zeggen over de mate van moeite met het interpreteren. In onderstaande tabel is de tijdsduur van de individuele artsen weergegeven en de gemiddeldes per versie.

Versie	Tijdsduur	Versie	Tijdsduur
Casus 1 – App	10.14 min	Casus 2 - App	04.51 min
	08.00 min		10.05 min
	06.30 min		11.31 min
	05.10 min		
	Gemiddeld: 7.29 min		Gemiddeld: 8.49 min
Casus 1 – EPD	15.21 min	Casus 2 - EPD	11.36 min
	06.10 min		09.17 min
	07.45 min		10.41 min
	05.49 min		
	Gemiddeld: 8.46 min		Gemiddeld: 10.31 min

Uit bovenstaande tabel blijkt dat bij casus 1 de artsen er 1.17 minuut korter over deden wanneer ze de app gebruikten dan wanneer ze het EPD gebruikten. Bij casus 2 deden de artsen er 1.42 minuut korter over met de app dan met het EPD. De artsen deden dus gemiddeld korter over een casus met gebruik van de app dan met gebruik van het EPD, ongeacht de casus dus ongeacht de ambiguïteit van de casus.

Een mogelijke verklaring hiervoor is gezocht door het bestuderen van de verschillen in de gedachteprocessen. Daarvoor is gekeken naar de operator ‘bestuderen’, om te zien of de artsen in het EPD meer bestuderen dan in het EPD. Dit bleek niet het geval, zo laat onderstaande tabel zien:

<i>Casus</i>	<i>Systeem</i>	<i>Aantal keer ‘bestuderen’</i>	<i>Systeem</i>	<i>Aantal keer ‘bestuderen’</i>
1	App	35	EPD	27
		23		15
		14		16
		19		15
	Gemiddeld: 22.75 = 23		Gemiddeld: 18.25 = 18	
2	App	10	EPD	21
		21		12
		24		20
	Gemiddeld: 18.33 = 18		Gemiddeld: 17.67 = 18	

Uit de tabel blijkt dat er weinig verschil zit in het aantal keren bestuderen. Per arts en per casus zijn de verschillen best groot, maar het gemiddelde is bij bijna elke casus en elk systeem hetzelfde. Alleen bij casus 1 met het gebruik van de app was het gemiddeld aantal keer bestuderen hoger dan bij de andere versies. Dit zou nog eerder ervoor pleiten dat het de artsen meer moeite kost wanneer ze met de app werken. Een verklaring daarvoor zou kunnen zijn dat de artsen nog moeten wennen aan de app en niet weten welke informatie ze in de app kunnen vinden. Bij de follow-up vragen werd echter vaak benoemd dat de artsen de app overzichtelijk en logisch gestructureerd vonden en dat ze een snel beeld kregen van hoe het met de patiënt gaat.

Kijken we echter naar wat en hoeveel er bestudeerd wordt, kan er wel een verklaring worden gevonden. In veel van de gevallen wordt er namelijk bij het gebruik van het EPD de anamnese van andere artsen gelezen alvorens andere waardes te bekijken of soms zelfs nog voor de samenvatting en situatieschets te lezen. Soms wordt ook later nog in het gedachteproces de anamneses gelezen. Dit is in verhouding tot de waardes en de kleuren die in de app worden weergegeven veel langere tekstuele informatie. Eén van de artsen leest ook de anamnese in het begin van het gedachteproces wanneer de app wordt gebruikt. Vijf van de artsen leest de anamnese in het begin wanneer ze het EPD gebruiken. Dit is bijvoorbeeld te zien in onderstaand voorbeeld. Vanaf de onderstreping is aangegeven dat de anamnese wordt bekeken:

“Ik ga even lezen hoor. Meisje van 3 jaar oud. Moet ik het gewoon hardop voorlezen? Meisje van 3 jaar oud met lokaal recidief ETMR (embryonal tumor with multilayered rosettes) links parietaal waarvoor resectie op 04-06. Weekend opgenomen met epilepsie waarvoor Kepra. Gezwollen brein waarvoor dexta door post operatief. Tijdens OK ondersteuning vulling en fenylefrine vanwege bloedverlies. Morfine onderhoud bij pijn starten. Dexamethason door. Neurocontroles a 1 uur. Situatie: Je hebt nachtdienst op de IC, het is 06.00 uur in de ochtend en je wordt wederom gevraagd door de verpleegkundige te komen kijken naar de zeer forse urineproductie de afgelopen uren. Oké. Ik zou eerst even kijken wat er al bekend is over dit meisje, waarom zij eventueel een forse urineproductie zou kunnen hebben. Dus ik ga heel even in de eerste consulten lezen wat er al bekend is over dit meisje. Gedetubeerd teruggekomen van

OK, post-operatief 450 cc bloedverlies, staat er. 1x 170 cc bloedtransfusie, komt gedetubeerd over van OK, geen ondersteuning. Hartfrequentie 102. Instabiele bloeddruk, nu geen ondersteuning nodig, neuro: brein gezwollen. Dexa gehad per operatief op verzoek van neurochirurgie. Temp niet gemeten per operatief. 2x kefsol gehad, hoeft niet gecontinueerd te worden. Gastro-intestinaal: mag een beetje drinken. En de neurochirurg schrijft op: *recraniotomie en verwijdering tumor vergroeid met de falx*” (A2).

Het lijkt er dus op dat de artsen wellicht standaard of automatisch de anamnese lezen wanneer ze een patiëntcasus openen in HiX. Dit zou een verklaring kunnen zijn voor dat ze langer deden over de casussen in het EPD dan de casussen in de applicatie.

Applicatie is intuïtief en heeft een logische structuur

Tijdens het gedachteproces en bij de follow-up vragen kwam naar voren dat de artsen de structuur of de opbouw van de applicatie logisch vonden. De indeling in organen leek de artsen te helpen in hun besluitvormingsproces. Dit komt overeen met de theorie van March (2010) over hoe mensen verhalen en modellen prettig vinden voor hun begrip. Het overzicht in de app geeft een eerste indruk van de patiënt, zo stellen zes van de artsen. Je hebt met een paar klikken een beetje een indruk. Door de overzichtelijkheid zorgt het er ook voor dat de artsen sneller een overzicht hebben van wat er aan de hand is met de patiënt, zo stellen vier van de artsen: *“Maar ik vond ook bij deze eerste casus wel overzichtelijk dat je meteen ziet waar liggen de alarmpunten zeg maar. En dat je nog steeds de ABCD structuur kan aanhouden. Dat is wel prettig vind ik”* (A5). Deze structuur van alarmpunten lijkt in de gedachteprocessen logisch voor de artsen te zijn. Zo wordt er bijvoorbeeld gezegd: *“Dan ga ik eerst even kijken naar het effect op andere organen”* (A1). Dit is precies wat in de applicatie gefaciliteerd wordt. Toch worden er ook tegenstrijdige reacties getoond die laten zien dat sommige artsen de structuur van HiX juist ook handig vinden:

“Nouja ik wil eigenlijk nog wel weten of er nog eerder lab. Wat er eerder aan lab is gedaan. Oh kijk dat staat hier makkelijker. Qua overzicht, of daar ooit al een keer een urine natrium bijvoorbeeld van is gedaan. Dat zou ik dan niet hier kunnen zien want dat staat bij, dit is allemaal bloed” (A2).

Deze arts was specifiek op zoek naar lab-uitslagen, dus het is dan ook niet verrassend dat een overzicht van alle lab-uitslagen als gemakkelijk worden gezien. Toch zijn er ook artsen die ondanks de kennis van de huidige systemen nog zoeken naar waar ze bepaalde informatie kunnen vinden: *“Heeft deze patiënt koorts? Weet ik dat? Dan zou dat in de metingen moeten zijn ergens. Of op het circulatieblad staat het ook”* (A3). De app zou hierin wellicht intuïtiever in kunnen zijn.

Bij de follow-upvragen werd door bijna elke arts aangegeven dat de structuur van de applicatie als intuïtief en gebruiksvriendelijk wordt ervaren:

“Waar je dus in HiX, ging ik denk ik meer eerst van anamnese lichamelijk onderzoek naar aanvullend onderzoek, naar. Dus dat is een volgorde die je meer gewend bent natuurlijk om in na te denken. Maar misschien dat het inderdaad wel efficiënter gaat op den duur als je verschillende resultaten in één keer bekijkt” (A5).

Kleuren van de organen in de applicatie

Uit het gedachteproces van de artsen vallen verschillende dingen op. Twee artsen noemen in het gedachteproces letterlijk iets over de kleuren. Ze geven hen aanwijzing om ernaar te kijken: *“Het valt me wel op dat die nieren geel zijn”* (A5) en *“Andere waardes maar even kijken, alles wat geel is aanklikken op de app”* (At1). Dit wordt bij de follow-upvragen door een arts bevestigd: *“Als je ernaar kijkt, dat je wel ziet welke een afwijkende kleur hebben, dus dan kan je zeggen dat is je orgaan van interesse. Dat is een orgaan waar je aandacht aan moet besteden”* (A3). Uit de gedachteprocessen bij

de artsen die het EPD gebruikten zou opgemaakt kunnen worden dat ze de ondersteuning van de kleuren hadden kunnen gebruiken om bepaalde waardes te duiden. Bijvoorbeeld bij het bepalen of de saturatiewaardes van het kind goed is. Daarvoor is het nodig om te weten welke waardes normaal zijn voor een kind van een bepaalde leeftijd. Bij het gebruik van het EPD probeert een arts daar zelf duiding aan te geven: *“Ik zit nog even naar de andere waarden te kijken van, de saturatie kijk dat is normaal zouden we zeggen dat dit laag is, maar bij een kind met zo'n hartaandoening kan het zijn dat het normaal is dus”* (A2). Een kleur had kunnen aangeven of deze waardes voor dit kind normaal zijn. De applicatie moet dit natuurlijk wel voor een individueel kind kunnen bepalen. Het is één ding dat het zou kunnen bepalen of waardes over het algemeen normaal zijn voor een kind van een bepaalde leeftijd, maar het moet ook nog bepaald worden voor een individu met een bepaald ziektebeeld. In dat geval is er dus sprake van ambiguïteit en is het de vraag of de app dit zou kunnen aangeven. Daarnaast wordt er vaak door de artsen bepaalde waardes afgegaan, waarna ze een conclusie daarover trekken:

“Natrium wat heel interessant is in deze casus, dat is heel stabiel. Dus dat pleit een beetje tegen een diabetes insipidus. Kalium is ook stabiel, trouwens dat natrium is wel, nee dat valt wel in de trend. Calcium stabiel, nierfunctie netjes, geen daling in Hb en Ht ook stabiel. Oké dus lab ziet er eigenlijk wel goed uit” (A1).

Deze arts werkte in het EPD. Dit laat zien dat de kleuren wellicht richting hadden kunnen geven hoe het met de patiënt gaat of wellicht twijfel kunnen wegnemen over hoe de waardes geïnterpreteerd moeten worden. Maar de vraag is ook of dat goed zou zijn. Nu wordt er extra nagedacht over of de natriumwaardes wel of niet in de trend vallen. Daarnaast is te zien dat alle waardes eerst af moeten worden gegaan voordat er een conclusie kan worden gegeven, terwijl met de kleuren een arts wellicht ook al met een ‘gekleurd’ beeld de data induikt. Dan hoeft het alleen te worden bevestigd en minder geïnterpreteerd te worden. Of het goed zou zijn dat de arts in een bepaalde richting wordt gestimuleerd, is de vraag.

Andere bevindingen uit het gedachteproces wijzen erop dat de kleuren juist geen richting geven. De artsen maken alsnog hun eigen interpretatie: *“Hartfrequentie, ze is 3 jaar, 89 nou dat is volgens mij normaal”* (A5). In principe had deze arts uit de kleur van het orgaan kunnen afleiden of de hartfrequentie afwijkend was of niet, maar daar is niets over opgemerkt. Uit de antwoorden van de follow-upvragen komt juist dat de artsen de kleuren behulpzaam vinden, maar ook dat ze normaalwaardes missen. Zo stellen twee artsen bij het zien van de casus in de app die ze oorspronkelijk in het EPD kregen, dat het hen had kunnen helpen in het beeld van de patiënt:

“Ja dat had mij geholpen. [...] 1. Je ziet meteen waar het punt zit, dat is namelijk rood en dat is als enige orgaan rood. 2. Je ziet dus op die andere punten ook dat dit de andere organen zijn, dit komt door de tachycardie omdat ie onrustig is en dat was mij minder goed, had ik wel gelezen in die cursus en had ik minder duidelijk geregistreerd omdat dit zo geel was geweest en ik dat het qua infectieus pad mij dat eerder was opgevallen dat die opgelopen temperaturen had en CRP. En daar moet je helemaal voor in HiX klikken en kijken wat hoe de oploop is etc. En wat ik hier dus fijn aan vind dus het zit er al wel in, dat wist ik niet, dat je dus de laatste positieve kweekuitslag krijgt. Ja, dit had mij geholpen” (A3).

Een andere arts merkt op bij het zien van de casus in de app dat er bepaalde informatie instaat die niet in het EPD stond (A5). De arts had die informatie dus niet in het EPD gevonden. De app had ervoor kunnen zorgen dat er een beter overzicht was van de informatie. Dit stelt ook één van de artsen: *“Ik denk dat je misschien minder snel dingen over het hoofd ziet”* (A5). Wat opvallend is, is dat drie artsen stellen de normaalwaardes te missen in de applicatie. In principe zouden de kleuren aangeven wat voor die patiënt goede of slechte waardes zijn. Ze willen zien waar de grens ligt en of iets heel erg afwijkt of een beetje afwijkt. Dat de kleuren kunnen helpen, zou kunnen zijn voor de volgorde van interesse: *“Het helpt misschien in de volgorde waardoor je iets afgaat”* (At1).

Bijna alle artsen zeggen bij het lezen van de situatieschets al een beeld in hun hoofd te hebben wat het zou kunnen zijn en op basis daarvan gaan zoeken naar informatie. Ze werden in die zin minder geleid door de kleuren. Wel stellen ze dat ze de rode en eventueel gele kleuren zouden aanklikken om te kijken wat daarmee aan de hand is. Eén van de artsen stelt ook te willen weten waarom groen groen is, en het daarom zou checken. Uit het experiment bleek echter dat lang niet alle artsen alle symbolen bekeken of het klopte. Dat zou in principe ook niet noodzakelijk zijn wanneer het niet een onderdeel vormt van het probleem: *“Ik denk dat ik sommige organen niet zou bekijken. Als ze niet onderdeel zijn van het probleem”* (A3). Drie van de artsen stellen dat ze ook de groene kleuren zouden bekijken wanneer die bij de situatieschets passen.

Bij het vragen naar wat de artsen van de kleuren vonden, wordt ook door drie van de artsen genoemd dat ze willen weten hoe de kleuren worden bepaald. Dit blijkt echter niet uit hun gedachteproces. Op geen enkel moment wordt genoemd dat ze kleuren afwijkend vinden of iets anders wat zou kunnen duiden op onbegrip of wantrouwen. Ze hebben dus in principe wel vertrouwen in dat de kleuren kloppen, maar vragen zich wel af waarop ze zijn gebaseerd:

“En ik zou nog wel willen weten voor gastro-intestinaal wat maakt waarom iets groen of oranje of rood is. Want als het alleen op een glucose-waarde is en wat er is ingevuld in Metavision aan voedingsstand zeg maar, ja dan daar zou ik dan niet op kunnen varen” (A3).

Hoewel ze dus achteraf zich afvragen op basis waarvan de kleuren worden bepaald en veranderen, hadden ze er in het gedachteproces geen moeite mee. Een ander punt dat wordt aangehaald is de artefacten die in de metingen kunnen zitten:

“Maar ik geloof zeker dat dit is wat er in de praktijk is gebeurd. Wat ik me wel afvraag, stel je voor dat er vier uur geleden precies een verpleegkundige bezig was met de lijn en dat er even een hoge CRP of iets gedetecteerd zou worden. Klopt die hier dan ook in en detecteert die dan als afwijkend of niet?” (A6).

Dit kan erop duiden dat er wordt afgevraagd of de kleuren gebaseerd zijn op één specifieke meting, of dat het een gemiddelde is van bijvoorbeeld de afgelopen twee uur. Deze discrepantie tussen het gedachteproces en de vragen achteraf is specifiek geanalyseerd met betrekking tot de trend. Veel van de respondenten noemden namelijk dat ze een trend missen in de applicatie, maar niets uit het gedachteproces wees erop dat ze veel moeite hadden met hoe het in de app is weergegeven. Met het klikken op het tijdspad in de app worden ook conclusies getrokken over of waardes toegenomen of gedaald zijn. Wellicht vinden ze het inderdaad wel prettiger zoals ze zelf zeggen, maar hebben ze het dus niet per se nodig om tot een conclusie van de trend te komen. Bij het gedachteproces wordt ook niet genoemd dat ze bepaalde waardes missen (bijvoorbeeld wat de waarde bij -1 uur was) en noemen ze ook bij het klikken op het tijdspad niet dat ze een trend missen. Daarnaast komen de artsen ongeveer gelijk verdeeld zowel tot een diagnose in de app als in het EPD. Het is soms lastig te zeggen met welke aanwijzingen ze precies hun diagnose vormen, maar het lijkt erop dat sommigen met enkel de informatie uit de app tot een diagnose kunnen komen, anderen vooral de informatie uit het EPD gebruiken en sommigen uit beide informatie halen en dan tot een beeld komen. Daarmee lijkt het in ieder geval niet zo te zijn dat ze per se een trend nodig hebben om tot een diagnose te komen.

Wat een verklaring zou kunnen zijn voor het missen van de trend, is dat het wellicht onduidelijk is of het tijdspad in de app bestaat uit losse punten uit de tijd (dat bij -2 uur de meting wordt weergegeven die precies op dat tijdstip wordt gedaan) of dat het bestaat uit gemiddeldes van de trend. Ook kan het zijn dat ze het beeld van de trend in één-oogopslag missen, waardoor je in één keer kan zien hoe het is verlopen over de tijd. Aan de andere kant is dit in principe ook zichtbaar aan de hand van de kleuren in de app.

Wennen aan de applicatie

Veel van de artsen benoemen dat ze nog wel zouden moeten wennen aan de app, omdat ze nu niet weten wat ze er allemaal in kunnen vinden en ze gewend zijn aan waar ze dingen kunnen vinden in HiX en Metavision. Dit kan het experiment hebben beïnvloed in de manier waarop de artsen omgingen met de app, bijvoorbeeld hoe vaak ze switchen naar het EPD om daar informatie te zoeken. Dat ze moeten wennen aan de applicatie bleek ook uit hun gedachteproces: *“Dan is de diurese niet beschreven in de app. Of staat het hierbij, nee”* (A1); *“Ik kijk nog even bij de andere dingetjes wat die betekenen”* (A2); *“Dan zou ik ten eerste willen weten, de vochtbalans. Is dit de vochtbalans? Ik ga gewoon maar even klikken”* (A3).

5.4 Verschillen in differentiaal diagnose

Om het interpretatieproces van de artsen te duiden, is ook gekeken naar de verschillen in de differentiaal diagnoses die ze uiteindelijk stelden. Wanneer de artsen met het gebruik van de app tot andere diagnoses zouden komen dan met gebruik van het EPD, zou dit kunnen duiden op een verschil in interpretatieproces. Hiervoor is ook de volgende hypothese opgesteld:

H2: De medicus komt tot een uitgebreidere differentiaal diagnose met gebruik van de app dan met gebruik van het EPD.

Verschillen diagnoses casus 1

Bij casus 1 werd zowel bij gebruik van de app als bij gebruik van het EPD door bijna elke arts de differentiaal diagnose virale infectie en hartfalen genoemd. Er waren twee artsen die enkel virale infectie benoemden, de een gebruikte de app en de ander gebruikte het EPD. Dit kan verklaard worden door de leeftijd en ervaring van deze artsen, zij zijn namelijk allebei (nog) niet in opleiding en rond de 25 jaar, de andere artsen zijn wel in opleiding en zijn boven de 30 jaar. Er kan dus gesteld worden dat er geen verschil is tussen de diagnoses die gesteld werden bij gebruik van de app en de diagnoses die gesteld werden bij gebruik van het EPD. Bij casus 1 komen de artsen dus niet tot een uitgebreidere differentiaal diagnose met het gebruik van de app dan met het gebruik van het EPD.

Verschillen diagnoses casus 2

Bij casus 2 is er net iets meer verschil in de gestelde diagnoses van de artsen tussen het gebruik van de app en het gebruik van het EPD. Bij het gebruik van de app stellen twee artsen dat het probleem hormonaal is. Eén arts noemt dat het probleem kan zijn dat het kind te veel gevuld is,³ of dat het hormonaal is. Twee van de drie artsen die de app gebruikten, noemden dus niet als optie dat het kind te veel gevuld zou kunnen zijn. Bij het gebruik van de app noemen twee van de drie artsen dit wel. De ander gaf aan dat het hormonaal is. De artsen die het EPD gebruikten, kwamen dus tot een uitgebreidere differentiaal diagnose dan de artsen die de app gebruikten. De twee jonge artsen met minder ervaring kwamen beide tot de twee opties, dus ervaring en leeftijd speelt hier geen rol. Wat wel een rol kan spelen is de ervaring in het vakgebied. Drie van de artsen noemen namelijk dat ze minder ervaring hebben met een dergelijke casus als casus 2, waarvan twee daarvan inderdaad enkel het probleem als hormonaal beschouwen. Daarvan kreeg één de casus in de app, de ander in het EPD. De derde arts die enkel hormonaal als probleem beschouwde, gaf aan casus 2 juist makkelijker te vinden, deels door het gebruik van de app. De artsen die stellen minder ervaring te hebben met een dergelijke casus, zijn dus gelijk verdeeld over de app en het EPD. Bij casus 2 kwamen de artsen met gebruik van het EPD dus tot een uitgebreidere differentiaal diagnose dan de artsen met gebruik van de app.

³ ‘Te veel gevuld’ betekent dat er bij de operatie veel vocht is toegediend waardoor de patiënt na de operatie veel vocht uitplast.

Over hypothese 3 kan dus gezegd worden dat medici niet tot een uitgebreidere differentiaal diagnose komen met gebruik van de app dan met gebruik van het EPD. Uit het experiment blijkt dat bij casus 2 de artsen zelfs tot een minder uitgebreide differentiaal diagnose komen met gebruik van de app dan met gebruik van het EPD. Wel is de vraag hoe geschikt de casus is, aangezien de helft van de artsen aangeven minder ervaring te hebben met een dergelijke casus. Hierdoor kan het zijn dat ze de juiste informatie niet herkennen die nodig is om tot een uitgebreidere differentiaal diagnose te komen. Daarnaast zou ook mee kunnen spelen dat de diagnose van 'te veel gevuld' een diagnose is waarbij er niets aan de hand is met de patiënt. Het zou kunnen dat mensen in een experiment extremere uitkomsten verwachten.

Bij de follow-upvragen wordt over het verschil tussen de casussen door drie van de artsen gezegd dat de app uitnodigt om alles te bekijken. Je kan makkelijk ergens op klikken om het even te checken, dus dat kan ervoor zorgen dat je minder snel dingen over het hoofd ziet. Dit zou ertoe bijdragen dat artsen tot een uitgebreidere differentiaal diagnose komen, maar dit was niet het geval. Dat hoeft niet te betekenen dat dit in de toekomst of bij andere casussen het niet zou kunnen helpen in het creëren van een overzichtelijk beeld van de patiënt en dingen minder snel over het hoofd zien.

5.5 Verschillen tussen de casussen met betrekking tot ambiguïteit

In het experiment zijn er twee casussen ontwikkeld die verschilden in mate van ambiguïteit. Casus 1 kon gezien worden als een complexere casus dan casus 2. Daarbij is gekeken of de mate van ambiguïteit een rol speelde in welke mate de applicatie kan bijdragen. Daarvoor is ook een hypothese opgesteld:

H3: Hoe hoger de mate van ambiguïteit van de casussen, hoe meer voordeel de app geeft voor de interpretatie van de data door de medicus ten opzichte van het EPD.

Voordeel = met minder moeite, sneller en uitgebreidere differentiaal diagnose

Hoewel casus 2 als minder ambigu werd bestempeld dan casus 1 vooraf, benoemden drie artsen dat ze casus 2 moeilijker vonden omdat ze er minder ervaring mee hadden. Twee van deze artsen beoordeelden de casus in het EPD, één deed dat in de app. Hoewel casus 1 is opgezet als een casus die meer ambigu is dan casus 2, bleek in het experiment dit voor sommige artsen juist andersom te zijn door hun ervaring met dergelijke casussen. In het bepalen van de ambiguïteit en dus complexiteit van de casus is deze ervaring niet meegenomen.

Minder moeite

Bij casus 2 werd vaker geswitcht naar het EPD dan bij casus 1. Bij casus 1 werd meer bestudeerd dan bij casus 2 bij gebruik van de applicatie. Bij gebruik van het EPD was dit gelijk. Verder was het lastig om noemenswaardige verschillen te ontdekken tussen casus 1 en 2 wat betreft de moeite die de artsen hadden met de interpretatie van de data.

Sneller

Bij casus 1 deden de artsen er 1.17 minuut korter over wanneer ze de app gebruikten dan wanneer ze het EPD gebruikten. Bij casus 2 deden de artsen er 1.42 minuut korter over met de app dan met het EPD. Wanneer casus 2 beschouwd wordt als meer ambigu dan casus 1 omdat de artsen daar minder ervaring mee hadden, zou gesteld kunnen worden dat ze er qua snelheid een klein voordeel van hadden.

Uitgebreidere differentiaal diagnose

Hoewel casus 2 dus als moeilijker werd beschouwd, gaf de applicatie daar niet per se meer voordeel mee. De artsen kwamen namelijk met het gebruik van het EPD tot een uitgebreidere differentiaal

diagnose dan de artsen die de applicatie gebruikten. Bij casus 1 was er geen verschil tussen het gebruik van de app en gebruik van het EPD wat betreft de uitgebreidheid van de differentiaal diagnoses.

Het is lastig om deze resultaten te bespreken omdat niet goed geduid kan worden of de mate van ambiguïteit zoals die in de casussen verwerkt is inderdaad een goede indicator kan zijn voor de mate van ambiguïteit.

5.6 Gebruik van de applicatie

De manier waarop de artsen het doel van de applicatie duiden en het vertrouwen erin zouden de uitkomsten van het experiment beter kunnen duiden. Wanneer opvattingen hierover erg zouden afwijken van wat er op dit moment beoogt wordt met de app, zouden de resultaten ook op een andere manier geïnterpreteerd kunnen worden.

Doel van de applicatie

Door vier van de artsen wordt gezegd dat ze de applicatie zouden zien als een soort alarmeringssysteem. *“Het is een soort monitor die slim kan nadenken. Maar alsnog moet je zelf als dokter blijven nadenken over wat er aan de hand is”* (A1). *“Bijvoorbeeld als je erbij wordt gevraagd bij een patiënt die acuut iets achteruitgaat. Persoonlijk ga ik altijd eerst zelf op de monitor kijken want dan heb je het meest recente beeld van het kind. Maar al sje dan wil terugkijken van hoe was het eerst en wat is er al bekend dan zou ik inderdaad dit gebruiken en zeker in de diensten, omdat je dan niet misschien de tijd hebt om eerst door alle waarden en statussen en dergelijke te lezen”* (A2). De artsen stellen ook dat het op bepaalde punten niet per se een toegevoegde waarde is: *“Dus dit helpt je volgens mij in het alarmeren puur op de getallen, maar eigenlijk dan nadenken wat is er nou, welke overwegingen zijn er, daar helpt dit natuurlijk nog niet echt bij want daarvoor heb ik toch echt HiX nodig om te kijken naar wat was de echo, hoe was het de afgelopen dagen”* (At1). De overwegingen van andere artsen wordt gemist. Daarnaast wordt de app ook gezien als iets dat in de dienst nuttig kan zijn, wanneer je ‘sneller dingen op een rijtje moet krijgen’ (A4) of als je erbij wordt gevraagd. Dan gaat het om situaties dat de arts het kind niet goed kent of niet eerder heeft gezien.

De applicatie zou volgens drie van de artsen niet als een derde systeem erbij moeten komen. Zij zien wel de toegevoegde waarde ervan, maar vragen zich af of ze het wel echt gaan gebruiken als het nog iets is waar ze mee kunnen werken, omdat er al zoveel is. Daarnaast wordt gezegd dat de overwegingen van de andere artsen worden gemist in de app en één van de artsen mist meer waardes uit bijvoorbeeld het lab.

Vertrouwen in de app

Veel van de artsen zeggen dat het vertrouwen in de app een kwestie van ervaring is. Na een tijdje weet je of de kleuren overeenkomen met de waardes die je ziet, waardoor het vertrouwen groeit. *“Maar ik zou niet meteen gaan rennen na elke rode long want misschien is het wel gewoon een meetfout. [...] Je blijft wel zelf verantwoordelijk voor wat je doet, maar je kan best een beetje geholpen worden en dat is wel handig hierin”* (A1). Er wordt ook veel gesteld dat ze de waardes zagen, en daarmee concludeerden of konden zien dat het klopte:

“Ik lees natuurlijk zelf ook wat er staat, ik bedoel ik klik erop maar dan ik zie oh 65 keer per minuut, oh ja dat is inderdaad te snel dus dat het, die kleurtjes en ook bij die blauwe kleurtjes die registreerde ik wel, maar ik heb wel zelf nog nagedacht is het ook inderdaad goed of niet goed? Dus ik had er wel vertrouwen in maar ik denk toch dat je kritisch blijft kijken” (At1).

Ook wordt genoemd dat wanneer de app gevalideerd en gecertificeerd is, ze het dan wel zouden vertrouwen. In dat geval vertrouwen ze dus op autoriteit.

5.7 Samenvattend

Met het experiment is getracht een antwoord te vinden op de vraag *Hoe wordt waarde toegekend aan informatie uit de app UrSmartStatus door medici?* In de analyse van het experiment is er gekeken naar de verschillen in het gedachteproces en besluit van de artsen tussen het gebruik van de app en het gebruik van het EPD voor twee casussen die verschilden in ambiguïteit. Om die verschillen te duiden zijn er drie hypothesen opgesteld die iets zeggen over de hoeveelheid moeite de interpretatie kostte voor de artsen, de uitgebreidheid van de differentiaal diagnose en op welke manier de app een eventueel voordeel kan geven voor casussen die in sterke mate ambigu zijn. Daarnaast kan er ook iets gezegd worden over hoe de artsen het gebruik van de applicatie zien en zullen de resultaten worden besproken in het licht van de overige verwachtingen uit het theoretisch kader. De conclusies die worden getrokken uit de resultaten moeten gezien worden als de best mogelijke conclusie die er getrokken kan worden op basis van de beschikbare data, maar beperkt zich wel tot dit specifieke experiment. Met het bevestigen of ontkrachten van de hypothesen geldt dit dan ook voor deze experimentele setting met het kleine aantal deelnemers en casussen. Het is daarmee de vraag of dezelfde conclusies kunnen worden getrokken wanneer het experiment wordt uitgevoerd met een groter aantal deelnemers en meerdere verschillende casussen. In die situatie kan meer gezegd worden over de werking van de applicatie. Wel kan het experiment hiervoor als blauwdruk dienen en laten de opzet en resultaten hiervan zien dat met een groter aantal deelnemers en casussen het experiment geschikt kan zijn om voor verdere tests te gebruiken.

Vershillen in gedachteproces

De artsen deden, ongeacht de casus, sneller over het komen tot een diagnose met gebruik van de app dan met gebruik van het EPD. De hoeveelheid informatie die de artsen observeerden, was bij casus 1 met gebruik van de app meer dan met gebruik van het EPD. Bij casus 2 observeerden de artsen minder informatie met gebruik van de app dan met gebruik van het EPD, maar dit verschil is kleiner dan bij casus 1. Hieruit zou opgemaakt kunnen worden dat bij casus 1 er met meer inzicht in de gezondheid van de patiënt een diagnose wordt gesteld met gebruik van de app dan met gebruik van het EPD. Bij casus 2 is dit andersom. Voor het stellen van die diagnose leek de app bij casus 1 net zo behulpzaam te zijn als het EPD, het was namelijk evenredig verdeeld over de artsen in welk systeem ze tot een besluit kwamen. Voor casus 2 leken de artsen meer informatie uit het EPD nodig te hebben. Er werd veel vaker geswitcht naar het EPD dan bij casus 1. De diagnose die ze stelden kwam voort uit zowel de app als het EPD. Bij casus 2 leek het EPD dus nog wel noodzakelijk om tot een besluit te komen. Dit zou verklaard kunnen worden door het gebrek aan ervaring met de casus en de artsen om die reden veel informatie nodig hadden om te kunnen achterhalen wat er aan de hand was. Uit de follow-upvragen bleek dat de artsen de app over het algemeen als intuïtief ervaren met een logisch overzichtelijke structuur. Uit de gedachteprocessen bleek dit soms wel, maar soms ook niet. Enkele artsen gingen mee in de structuur van de applicatie (*“Dan ga ik eerst even kijken naar het effect op andere organen”*), anderen vonden juist de structuur van bijvoorbeeld de lab-uitslagen ook heel prettig omdat ze specifiek zochten naar de lab-uitslagen. Er werd dan ook door sommigen direct geswitcht naar het EPD wanneer ze specifiek ergens naar op zoek waren. Dit zou verklaard kunnen worden door de gewenning aan het EPD.

De kleuren van de organen in de applicatie vielen bij sommige van de artsen op en deden daar ook wat mee (*“Het valt me wel op dat die nieren geel zijn”*), anderen viel het niet op en trokken zelf conclusies uit de waardes. Uit de follow-upvragen bleek dat de artsen de kleuren nuttig en behulpzaam vinden, maar er wordt ook veel gesteld dat ze de normaalwaardes missen om te kunnen zien of de gemeten waardes afwijken van wat normaal is. Dit lijkt tegenstrijdig, omdat de kleuren ook een indicatie geven

van hoever de waarde afwijkt van wat normaal is. Sommige van de artsen vragen zich af hoe de kleuren bepaald worden, op welk moment iets rood wordt en welke indicatoren daarbij horen. In het gedachteproces kwam dit echter niet naar voren. Er was geen enkel moment waarop er iets werd gezegd over de onjuistheid van de kleuren of op een manier opvallend te vinden. Dit verschil tussen het gedachteproces en de antwoorden bij de vragen achteraf kwam ook naar voren bij het missen van een trend. Bij de follow-upvragen werd vaak genoemd dat er een trend werd gemist, maar tijdens het gedachteproces kwam dit niet naar voren. Er werden conclusies getrokken over de veranderingen over tijd zonder iets te zeggen over het missen van een trend.

Hypothese 1: De medicus kan met minder moeite een interpretatie geven aan de data met het gebruik van de app dan met het gebruik van het EPD kan niet met volle overtuiging bevestigd worden, maar het tegenovergestelde kan in ieder geval ontkracht worden. De artsen leken in geen geval meer moeite met de interpretatie te hebben met gebruik van de applicatie dan met gebruik van het EPD. Voor casus 1 zouden de resultaten erop kunnen duiden dat met gebruik van de app de artsen minder moeite hebben met de interpretatie dan met gebruik van het EPD, omdat ze er sneller over doen, meer inzicht in informatie hebben, weinig switchten naar het EPD en ze zelf aangeven de app intuïtief en overzichtelijk te vinden. Voor casus 2 zijn de resultaten wat moeilijker te interpreteren. Met gebruik van de app deden de artsen er sneller over, maar switchten wel vaak naar het EPD en vonden minder informatie in de app dan in het EPD. Het gebrek aan ervaring met dergelijke casussen door de artsen kan een mogelijke verklaring zijn voor dit verschil tussen de casussen.

Verschillen in besluit

Voor casus 1 waren de differentiaal diagnoses die de artsen stelden hetzelfde bij gebruik van de app als bij gebruik van het EPD. Voor casus 2 werd er met gebruik van het EPD uitgebreidere differentiaal diagnoses gesteld dan met gebruik van de app. Dit volgt logisch uit de bevinding dat er bij casus 2 minder informatie wordt gevonden in de app dan in het EPD; dit is andersom het geval bij casus 1. Voor casus 2 kan bij dit resultaat meegenomen worden dat de artsen minder ervaring hadden met een dergelijke casus en daarom wellicht meer informatie uit het EPD nodig hadden.

Hypothese 2: De medicus komt tot een uitgebreidere differentiaal diagnose met gebruik van de app dan met gebruik van het EPD kan worden ontkracht. Bij casus 1 komen de artsen tot dezelfde differentiaal diagnose, bij casus 2 komen ze tot een minder uitgebreide differentiaal diagnose. Hoewel de artsen bij casus 1 tot dezelfde diagnose komen, hebben ze wel meer informatie tot zich genomen bij de app dan bij het EPD. Dit leidt dus niet tot verschillen in de diagnose, maar wellicht wel tot meer inzicht in de gezondheid van de patiënt. Dat de artsen bij casus 2 tot een minder uitgebreide differentiaal diagnose kwamen met gebruik van de app dan met gebruik van het EPD, zou eventueel te verklaren zijn uit hun gebrek aan ervaring. De diagnose die minder vaak werd genoemd bij gebruik van de app dan met gebruik van het EPD, is een diagnose die geen gevolgen heeft voor de patiënt. Dit is een verschijnsel dat wellicht minder voor de hand ligt wanneer men voor een experiment wordt gevraagd om een diagnose te stellen.

Mate van ambiguïteit

Bij de analyse is ook stilgestaan bij het verschil in mate van ambiguïteit van de twee casussen. Daarvoor is gekeken of het voordeel dat de app kan geven voor de interpretatie van de app afhankelijk is van de mate van ambiguïteit. Casus 1 was vooraf bepaald een complexere en daardoor een meer ambigue casus dan casus 2. Dat de artsen minder ervaring bij casus 2 hadden, zou voor de complexiteit in principe niet een verschil moeten maken. De mate van voordeel is geduïd als dat de artsen minder moeite hoeven te doen voor de interpretatie, sneller zijn en tot een uitgebreidere differentiaal diagnose komen. Zoals bij de eerste hypothese geconcludeerd, lijken de resultaten uit te wijzen dat de artsen bij casus 1 inderdaad

minder moeite hebben met de interpretatie van de data met gebruik van de app dan met gebruik van het EPD. Ook kwamen ze sneller tot een diagnose. De differentiaal diagnose was gelijk voor het gebruik van de app dan het gebruik van het EPD. Wel hebben ze meer informatie tot zich genomen in de app. Bij casus 2 is de moeite voor het interpreteren niet minder bij gebruik van de app dan bij gebruik van het EPD (maar ook niet per se meer), kwamen de artsen sneller tot een diagnose maar iets minder snel in vergelijking met casus 1 en was hun differentiaal diagnose minder uitgebreid bij gebruik van de app dan bij gebruik van het EPD.

Hypothese 3: *Hoe hoger de mate van ambiguïteit van de casussen, hoe meer voordeel de app geeft voor de interpretatie van de data door de medicus ten opzichte van het EPD* kan dus bevestigd worden. Het gebruik van de app lijkt bij casus 1 voordeliger te zijn voor de interpretatie van de arts dan bij casus 2.

Gebruik van applicatie

Het theoretisch kader gaf de verwachting dat de app nog niet naadloos zou aansluiten bij het werkproces van de artsen, doordat ze niet gewend zijn aan het werken met een dergelijk systeem. Uit de follow-up vragen komt dit ook naar voren. De artsen geven aan dat ze moesten wennen aan het gebruik van de applicatie. Ze wisten niet wat ze erin konden vinden en het nadenken in een andere structuur was ook wennen. Uit de gedachteprocessen leek dit ook het geval: vaak werd geswitcht naar het EPD om bepaalde informatie op te zoeken, terwijl deze ook in de app te vinden was. Het vertrouwen in de app lijkt geen probleem te vormen voor het gebruik. De artsen geven aan een dergelijke app wel te vertrouwen wanneer ze er meer ervaring mee hebben en zien dat de waardes kloppen en het een gecertificeerde applicatie is. De app zou volgens de artsen gezien kunnen worden als een alarmeringssysteem en handig voor een snel overzicht wanneer ze erbij worden geroepen. De meeste artsen zouden het wel als een toegevoegde waarde zien, maar vragen zich wel af hoe het zou bestaan naast alle andere systemen die er al bestaan.

6. Conclusie

Met een toenemende complexiteit van informatie in de vorm van big data en focus op evidence-based wetenschappelijke informatie in de medische wereld wordt er gezocht naar manieren om daar mee om te gaan (Murdoch & Detsky, 2013; Sim, 2016). Nieuwe technologieën als decision-support systemen met algoritmes doen hun intrede om artsen te kunnen helpen in hun besluitvorming en omgang met deze grote hoeveelheden aan informatie. In het UMC Utrecht is er een decision-support systeem in ontwikkeling: UrSmartStatus. Deze applicatie heeft als doel het ‘communicatieprobleem’ tussen de data en de arts op te lossen. Het structureert de data per tractus in plaats van inputgericht en geeft hier een waarde-oordeel aan door de organen een kleur te geven die aangeeft hoe goed of slecht het met het orgaan gaat, van groen, geel, oranje naar rood. De data die hiervoor gebruikt wordt zijn historische patiëntgevallen waaraan de status van huidige patiënt wordt gespiegeld. Hierbij is niet alleen de ambitie om de huidige status met de geschiedenis van de patiënt te laten zien, maar ook een prognose te geven van de gezondheid van de patiënt. Met de verandering van de manier van presenteren van en het toevoegen van waarde aan de medische data verandert de manier waarop de arts de data en informatie tot zich neemt. Er vindt een verandering in het waarde-creatie proces plaats, waarin de ontwikkelaars waarde toevoegen aan de data en de arts deze gaat interpreteren. In dit onderzoek is dan ook bestudeerd hoe dit proces van waarde-creatie eruitziet en verschillend is ten opzichte van de huidige situatie. Daartoe is de volgende hoofdvraag opgesteld:

Hoe krijgt medische data waarde in de decision-support applicatie UrSmartStatus in het UMC Utrecht?

Deze vraag is tweeledig: aan de ene kant voegen de ontwikkelaars waarde toe aan de medische data en aan de andere kant geven artsen een interpretatie aan de informatie uit de applicatie. In dit onderzoek is dan ook het waarde-creatieproces van deze twee groepen onderzocht. Daartoe werden bij de ontwikkelaars interviews gehouden waarbij drie grote onderwerpen werden besproken: het verschil tussen de applicatie en het EPD, besluitvorming en interpretatie van data. Bij de artsen werd er een experiment gedaan waarin de artsen twee verschillende patiëntcasussen kregen aangeboden, de één in de applicatie en de ander in het EPD. Hierbij spraken ze hun gedachteproces hardop uit om tot een diagnose te komen en kregen ze enkele vragen achteraf. Op deze manier kon het hele proces van ontwikkeling tot gebruik in kaart worden gebracht.

Waarde-creatie van de ontwikkelaars

Hoe de ontwikkelaars waarde toekennen aan medische data met de app UrSmartStatus, is te beantwoorden aan de hand van de drie thema's het verschil tussen de app en het epd, besluitvorming en interpretatie van data. Bij het verschil tussen de app en het epd speelt de ambiguïteit van presentatie een grote rol (March, 1987). Met de indeling per orgaan is de dat logischer gestructureerd waarbij er conclusies worden verbonden aan de data door middel van de kleuren van de organen. Deze zijn gebaseerd op de data van historische patiënten. Op die manier geven de ontwikkelaars of het algoritme een interpretatie aan de data door aan te geven hoe goed of slecht het gaat met een orgaan. De data krijgt op die manier niet alleen een waarde van ernst, maar ook een prioritering van volgorde van aandacht. Dit verandert ook de manier waarop de data binnenkomt bij de arts. Het is een meer top-down gerichte aanpak waarbij de arts eerst een globale indruk krijgt van de gezondheid van de patiënt en vervolgens gedetailleerder op de specifieke data kan ingaan. Dit is tegenovergesteld aan hoe de situatie nu is: artsen krijgen alle specifiek data en moeten daaruit zelf een oordeel vormen over de gezondheid van de patiënt. De ontwikkelaars creëren dus op twee manieren waarde. Enerzijds door de data op een andere manier te presenteren waardoor het werkproces van de arts meer centraal komt te staan (per tractus in plaats van inputgericht), anderzijds door een oordeel te geven aan de data door deze te vergelijken met historische patiënten en zodoende een ernst-interpretatie te geven met prioritering. Deze interpretatie is

volgens de theorie van March (1987). Door informatie ambigu te presenteren blijft er genoeg ruimte over voor een veelheid aan interpretaties. Op die manier zouden de kleuren niet zozeer leidraad vormen in het oordeel van de arts, maar blijft de arts zijn of haar kritische blik gebruiken om tot een oordeel te komen.

Voor deze interpretatie van de data wordt door sommige van de ontwikkelaars gesteld dat de ambitie is dat deze uiteindelijk accurater zal zijn dan die van de arts. Voor het rationele gedeelte van de besluitvorming zou dit in principe ook mogelijk moeten zijn: een algoritme kent geen subjectieve ervaring en emotie dus de vergelijking met andere patiënten kan als objectiever worden gezien. Hierbij is het wel de vraag wat 'objectiever' in dat opzicht betekent: bij het maken van algoritmes worden keuzes gemaakt die altijd bepaalde waarden zullen reflecteren van de ontwikkelaars (Mittelstadt et al., 2010). Wat de ontwikkelaars ook noemen is dat de klinische blik van de arts noodzakelijk blijft voor het komen tot een oordeel of besluit. Er moet 'genoeg' ruimte overblijven voor de interpretatie van de arts. Tegelijkertijd moet de arts wel zodanig veel vertrouwen hebben in de app om het te gebruiken. De applicatie beoogt dan ook om de computer te laten doen waar het goed in is (het rationele gedeelte van de besluitvorming) en de mens te laten doen waar het goed in is: de kritische, klinische logica. Uit deze bevindingen rezen vragen op voor het experiment met de artsen. Wordt de arts inderdaad geholpen in het rationele gedeelte van de besluitvorming, blijft er 'genoeg' ruimte over voor de interpretatie van de arts, hoeveel waarde kan er worden gehecht aan de kleuren wanneer deze zijn gebaseerd op het algoritme met rationale logica en in hoeverre wordt de arts beïnvloedt door de waarde die aan de medische data is gegeven en de richting die de prioritering van data kan geven in het eigen oordeel?

Waarde-creatie van de artsen

Hoe de artsen waarde toekennen aan de medische informatie uit de app UrSmartStatus, is te beantwoorden aan de hand van de vergelijking tussen het gedachteproces en besluit van de artsen bij gebruik van de applicatie en bij gebruik van het EPD. Uit het experiment lijkt geconcludeerd te kunnen worden dat de artsen niet meer moeite met de interpretatie hadden met gebruik van de applicatie dan met gebruik van het EPD. Dit verschilde voor de twee casussen, waarbij casus 1 vooropgesteld meer ambiguïteit bevatte dan casus 2. Bij casus 1 observeerden de artsen meer informatie in de applicatie dan in het EPD en konden daardoor meer inzicht hebben in de gezondheid van de patiënt alvorens ze een diagnose stelden. Ze deden sneller over het stellen van een diagnose, switchten weinig naar het EPD en vonden de app intuïtief en overzichtelijk. Voor casus 1 lijkt het er dan ook op dat de applicatie ervoor zorgt dat de artsen minder moeite hebben met het interpreteren van de data. Voor casus 2 is dit was lastiger te stellen. Daarbij waren de artsen sneller met de app dan met het EPD, maar switchten ze vaak naar EPD en vonden meer informatie in het EPD dan in de app. De artsen gaven aan minder ervaring met een dergelijke casus als casus 2 te hebben, wat een mogelijke verklaring kan zijn voor dit verschil tussen de casussen. Bij het stellen van de uiteindelijke diagnose lijkt de app geen verschil te maken of te leiden tot een minder uitgebreide differentiaal diagnose. Bij casus 1 maakt het namelijk geen verschil, maar bij casus 2 komen de artsen tot een minder uitgebreide differentiaal diagnose bij gebruik van de app dan met gebruik van het EPD. Dit verschil kan mogelijk ook verklaard worden door het gebrek aan ervaring die de artsen hadden met een dergelijke casus. De diagnose die ontbrak was een verschijnsel dat wellicht minder voor de hand ligt wanneer men als arts er wordt bijgevraagd. De verwachting zou kunnen zijn dat er iets aan de hand is wat gevolgen kan hebben voor het kind, maar deze diagnose zou dat niet hebben. Hoewel de artsen tot een minder uitgebreide differentiaal diagnose kwamen, vergaarden ze over het algemeen wel meer informatie uit de applicatie dan uit het EPD. Hiervan zou gesteld kunnen worden dat ze daardoor meer inzicht hebben in de gezondheid van de patiënt bij de app dan bij het EPD.

Het gebruik van de app lijkt in ieder geval voordeliger te zijn dan het EPD voor casus 1. Dit zou erop kunnen wijzen dat hoe meer ambigu de casus is, hoe voordeliger de app is ten opzichte van het EPD. Dit is echter wel met de kanttekening dat (een deel van) de artsen minder ervaring hadden met een casus als casus 2 en daardoor de resultaten vertekend kunnen zijn.

Vergelijking waarde-creatie van ontwikkelaars en interpretatie van artsen

De onderzoeksvraag *Hoe krijgt medische data waarde in de decision-support applicatie UrSmartStatus in het UMC Utrecht?* kan dus als volgt worden beantwoord. De waarde-creatie van de ontwikkelaars was tweeledig: de manier van presenteren en het geven van een waarde-oordeel aan de data door middel van de kleuren aan de hand van de vergelijking met historische patiëntgevallen. Door informatie ambigu te presenteren zou er genoeg ruimte moeten overblijven voor een veelheid aan interpretaties. Op die manier zouden de kleuren niet zozeer leidraad vormen in het oordeel van de arts, maar blijft de arts zijn of haar kritische blik gebruiken om tot een oordeel te komen. Dit komt eigenlijk ook voort uit het experiment. Hoewel de artsen benoemden de kleuren nuttig te vinden, gingen ze er niet (volledig) op af om hun oordeel te vormen. Zij gebruikten het soms om te kijken waarom er een mate van ernst werd weergegeven (geel, rood), maar gebruikten hun eigen logica om de informatie te bekijken en er conclusies uit te trekken. Het leek voor casus 1 te helpen om meer inzicht te vergaren over de toestand van de patiënt: ze observeerden meer informatie. Daaruit zou een voorzichtige conclusie getrokken kunnen worden dat bij een meer ambigu casus de app meer voordeel oplevert. In dat opzicht lijkt het erop dat de applicatie de arts inderdaad kan helpen in de rationele logica door aan te sluiten bij hoe ze denken, maar dat de klinische logica bij de arts blijft. Het zorgt er in ieder geval niet voor dat de artsen meer moeite hebben met de interpretatie.

Hierbij is het wel de vraag waar dit eventuele voordeel uit wordt gehaald. Hebben de artsen baat bij de structuur van de applicatie of zijn het de kleuren en daarmee het waarde-oordeel wat aan de data wordt gegeven? Het is lastig hier een sluitend antwoord op te geven. Uit het experiment lijkt het een combinatie te zijn van beide. De applicatie wordt als intuïtief en overzichtelijk beschouwd. Over de kleuren wordt gezegd dat ze nuttig zijn, maar uit het gedachteproces blijkt ook dat de artsen zelf hun oordeel geven door zelf conclusies te trekken uit de data. De kleuren lijken soms behulpzaam om de organen van interesse onder de aandacht te brengen, maar soms blijkt ook dat de kleuren voor de artsen niet opvallen. Uit de gedachteprocessen en vragen achteraf blijkt dat de artsen vooral vanuit hun eigen eerste gedachte over de patiënt de data gaan bekijken, en daarmee hun eigen volgorde afdelen. Bij bijvoorbeeld casus 1 werd door sommige artsen als eerste de rode longen aangeklikt, maar sommigen begonnen ook bij het gele hart. Ook worden de groene organen bekeken die passen bij het probleem, om een volledig beeld te krijgen van de patiënt. Bij de vragen achteraf vroegen sommige artsen zich ook af hoe de kleuren tot stand komen; wanneer is iets rood? De kleuren lijken dus niet heel sterk een voordeel te geven. De structuur zou daardoor wellicht meer helpen bij een overzicht creëren van de gezondheid van de patiënt. Uiteraard moet er rekening mee worden gehouden dat er altijd onbewuste processen bij de artsen die niet worden uitgesproken, waardoor er effecten kunnen zijn die niet geregistreerd zijn.

Uit bovenstaande kan dus geconcludeerd worden dat de ontwikkelaars waarde toevoegen door een waarde-oordeel te geven aan de data en deze ambigu te presenteren. Een voorzichtige conclusie uit het experiment is dat voor de minder ambigu casus de app minder tot geen voordeel lijkt te hebben voor de interpretatie voor de arts, maar heeft als kanttekening dat de artsen minder ervaring hadden met een dergelijke casus. Bij de ambigu casus lijkt de app wel een voordeel op te leveren voor de artsen: ze hebben minder moeite met de interpretatie maar komen tot dezelfde diagnose als bij gebruik van het EPD. Dit effect lijkt meer te danken aan de presentatie van de data (top-down en ingedeeld per tractus) dan aan het waarde-oordeel via het kleuren van de organen.

7. Discussie

Tijdens en na het onderzoek deden zich punten van discussie voor die hier zullen worden besproken. Deze hebben betrekking op de methodologie en de uitkomsten van het onderzoek. Daarnaast zullen er suggesties worden gedaan voor vervolgonderzoek.

Reflectie op methodologie

Bij het ontwerpen van het onderzoek leken uit het theoretisch kader ambiguïteit en onzekerheid van data twee belangrijke en interessante onderwerpen om te zien hoe die in het waarde-creatieproces tot uiting zouden komen bij de ontwikkelaars en artsen. Bij de interviews met de ontwikkelaars bleek onzekerheid echter een minder groot issue dan verwacht, omdat dit een principe is waar een computer goed mee om kan gaan. Tevens bleek bij het ontwerpen van het experiment dat het opnemen van het prognose-gedeelte van de applicatie in het ontwerp geen goede gelijke vergelijking zou opleveren met het EPD. In het EPD wordt er namelijk geen prognose gegeven. Dit onderdeel, en daarmee (een groot deel van) onzekerheid van data, is daarom niet in het experiment opgenomen. Er kan dus ook niets gezegd worden over hoe artsen de informatie over onzekerheid van data interpreteren.

Over de interviews met de ontwikkelaars kan opgemerkt worden dat de ontwikkeling van de applicatie nog in een heel vroeg stadium is. Zodoende kan de app zich dan ook nog heel anders ontwikkelen en kunnen tevens de ontwikkelaars andere inzichten ontwikkelen waardoor hun waarde-creatie wellicht anders zou uitpakken. Daarnaast was het voor sommige ontwikkelaars lastig om iets te zeggen over onderwerpen als ambiguïteit en onzekerheid van data, omdat zij niet als datawetenschapper aan het project werken. Zij focussen zich bijvoorbeeld meer op het design van de applicatie. Ook was ambiguïteit van data voor de ontwikkelaars nog geen onderwerp waar ze mee bezig zijn, dus wellicht dat de resultaten anders zouden uitpakken als zij hier verder in gevorderd zijn.

Voor het experiment zijn een aantal dingen op te merken. Zoals ook aangegeven in het methode hoofdstuk kon bij het hardop uitspreken niet het gehele gedachteproces gevolgd worden. Mensen lezen sneller dan dat ze het uit kunnen spreken en maken toch kleine notities in hun hoofd die ze niet uitspreken. Eén van de artsen merkte dit ook op bij de vragen achteraf. Hierdoor kunnen er effecten zijn van onbewuste processen die niet geregistreerd zijn. Het is lastig dit nog te optimaliseren. Wat wel geoptimaliseerd zou kunnen worden, is het aantal deelnemers en de variëteit aan casussen in het experiment. Meer deelnemers en meer verschillende soorten casussen zou het experiment betrouwbaarder maken. Omwille van het tijdsbestek van dit onderzoek was dit helaas niet mogelijk. Om die reden kunnen er ook geen sluitende uitspraken gedaan worden over de resultaten van het experiment. Een ander punt voor discussie is de manier waarop de casussen zijn verwerkt in de applicatie. Omdat de applicatie nog in ontwikkeling is, kon voor dit experiment de kleuren niet gebaseerd worden op historische patiëntgevallen. De kleuren zijn bepaald op intuïtie van de artsen verbonden aan het project. Om die reden is de vergelijking tussen de waarde-creatie van de ontwikkelaars en de interpretatie van de artsen niet geheel precies. Dit laat deels zien dat artsen wel een soort gedeelde intuïtie hebben (niemand vond de gekozen kleuren opvallend), maar zegt nog niets over hoe het systeem wat betreft de kleurbeoordeling straks zal werken.

De generaliseerbaarheid van het onderzoek is beperkt door de unieke case met beperkt aantal respondenten en deelnemers aan het onderzoek. De uitkomsten kunnen niet één op één vertaald worden naar andere onderzoekscontexten. De opzet van het onderzoek zou echter wel gebruikt kunnen worden voor andere onderzoeken naar de werking en implicaties van decision-support systemen. Het experiment bleek een goede manier te zijn om het interpretatieproces van de artsen in beeld te brengen en de verschillen te kunnen duiden tussen het gebruik van de nieuwe applicatie en het huidige EPD systeem.

Reflectie op onderzoeksbevindingen

Wat uit het onderzoek niet goed heeft kunnen blijken, is hoe er moet worden omgegaan met ambiguïteit van medische data in de applicatie. Rationele logica en klinische besluitvorming zijn verweven met elkaar en zijn niet op elkaar volgend. Met het verbeteren van de rationele logica middels de algoritmes in de applicatie moet deze nieuwe vorm van informatie op de juiste manier terechtkomen bij de arts. Met de kleuren van de organen wordt er een waarde-oordeel gegeven aan de data: er wordt aangegeven in hoeverre een patiënt afwijkt van historische patiëntgevallen en in welke mate van ernst. Uit het theoretisch kader lijkt op te maken dat ambiguïteit een rol speelt in die data. Het zou aan de arts zijn om deze ambiguïteit te duiden en de klinische interpretatie aan de situatie toe te voegen. In dit onderzoek kon niet blijken of de interpretatie van de arts 'achteraf' voldoende zou zijn om de data goed te kunnen duiden en daarmee of de kleuren een nuttige eerste indruk kunnen geven van de patiënt. Door de ontwikkelaars wordt genoemd dat een dynamische vorm van de applicatie een oplossing zou kunnen zijn voor dit eventuele probleem. De arts zou aanpassingen kunnen maken in de applicatie. Maar is het begrip van de ambiguïteit van de situatie ook niet nodig voor bepaling van de kleuren? Daarnaast is de gewenning van het EPD een factor om mee te wegen in de bevindingen. Artsen zijn gewend om op een bepaalde manier naar medische data te kijken en deze te behandelen. Uit de gedachteprocessen bleek dit ook zo te zijn: artsen hadden vaak een bepaald beeld in hun hoofd en zochten vervolgens naar informatie om dit te kunnen bevestigen of uit te zoeken. Ook zijn ze gewend om te denken in de structuur van het EPD. Een nieuwe manier van informatiepresentatie zal tijd moeten hebben om gewoonte te worden. Het toevoegen van de kleuren zou er wel voor kunnen zorgen dat artsen gealarmeerd worden wanneer er organen oranje of rood kleuren wanneer dit niet in hun beeld past. Dit kan ervoor zorgen dat ze een vollediger beeld hebben van de patiënt en kan wellicht voorkomen dat er informatie over het hoofd wordt gezien. Een 'gevaar' ervan zou zijn dat artsen eerder getriggerd worden tot overbehandeling. Bij lichtgroene of lichtgele organen mist er informatie. Dat zou kunnen aanzetten tot het volledig maken van de informatie.

Aanbevelingen voor vervolgonderzoek

Voor vervolgonderzoek is het aan te raden de artsen van tevoren kennis te laten maken met de applicatie en haar functies, zodat er minder sprake hoeft te zijn van onbekendheid van informatie. In het experiment viel namelijk op dat artsen vaak ook dingen in het EPD opzochten omdat ze wisten waar het daar stond en ze niet wisten of dit ook in de app zou staan. Dit zou het gebruik van de applicatie realistischer maken. In eerste instantie was het in dit onderzoek ook de bedoeling dat de artsen eerst een voorbeeldcasus zouden krijgen waarbij ze de status rustig konden bekijken om te wennen aan de functies. Dit was omwille van de tijd niet haalbaar. Hoewel dat wellicht had geholpen de resultaten betrouwbaarder te maken, zou ook dit wellicht niet voldoende zijn. De arts moet goed weten welke informatie er wel en niet te vinden is in de applicatie om het experiment goed te kunnen doen. Daarnaast vonden drie van de artsen de situatieschets onrealistisch. Een casus voor je krijgen zonder enige voorkennis en zonder het kind te kunnen bekijken vonden zij lastig om in te leven. Sommige artsen merkten ook op dat ze het kind zouden bekijken en beoordelen. Dit is toch een belangrijk onderdeel van het stellen van de diagnose. Dit zou in een volgend onderzoek meegenomen kunnen worden, om de situatieschets zo te maken dat het direct duidelijk is op welk moment de arts erbij wordt geroepen en wellicht een korte omschrijving van het lichamelijk onderzoek van de patiënt. Bij het experiment is het ook de vraag wat de arts kan helpen in de interpretatie: is dat de nieuwe manier van presentatie van data of is dat het waarde-oordeel uit de vergelijking met historische patiënten? Om dit goed uit te zoeken is het wellicht aan te raden om deze twee vernieuwingen apart van elkaar te testen. Ten slotte zijn de reflecties op de methodologie ook punten om mee te nemen voor vervolgonderzoek.

8. Literatuur

- Beerends, S. (2019, 9 maart). Smeltkroezen met een pottenbakrobot. *Setup*. Geraadpleegd via: <https://www.setup.nl/magazine/2019/03/smeltkroezen-met-een-pottenbakrobot>
- Berger, L., Bleichrodt, H., & Eeckhoudt, L. (2013). Treatment decisions under ambiguity. *Journal of Health Economics*, 32, 559-569.
- Berner, E. (Red.). (2007). *Clinical Decision Support Systems* (2e ed.). New York: Springer Science+Business Media.
- Brand, P. (2012). Interpretatie en oordeelsvorming beïnvloeden evidencebased richtlijnen. Commentaar. *Ned Tijdschr Geneeskd*, 156, 1-2.
- Bryman, A. (2012). *Social research methods* (4th edition). Oxford University Press.
- Cohen, G., Lynch, H.F., Vayena, E., & Gasser, U. (Red.). (2018). *Big Data, Health Law, and Bioethics*. Cambridge: Cambridge University Press.
- Cresswell, K., Majeed, A., Bates, D., & Sheikh, A. (2012). Computerised decision support systems for healthcare professionals: an interpretative review. *Informatics in Primary Care*, 20, 115-128.
- Ericsson, K.A., & Simon, H.A. (1984). *Protocol Analysis*. Cambridge: MIT Press.
- Esmailzadeh, P., Sambasivan, M., Kumar, N., & Nezakati, H. (2015). Adoption of clinical decision support systems in a developing country: Antecedents and outcomes of physician's threat to perceived professional autonomy. *International Journal of Medical Informatics*, 84, 548-560.
- Fonteyn, M.E., Kuipers, B., & Grobe, S.J. (1993). A Description of Think Aloud Method and Protocol Analysis. *Qualitative Health Research*, 3(4), 430-441.
- Garg, A., Adhikari, N., McDonald, H., Rosas-Arellano, P., Devereaux, P. J., Beyene, J., ... Haynes, B. (2005). Effects of Computerized Clinical Decision Support Systems on Practitioner Performance and Patient Outcomes: A Systematic Review. *Journal of the American Medical Association*, 293(10), 1223-1238.
- Geuijen, K., Moore, M., Cederquist, A., Ronning, R., & Van Twist, M. (2017). Creating Public Value in Global Wicked Problems. *Public Management Review*, 19(5), 621-639.
- Goodman, K. (2009). *Ethics, Computing, and Medicine: Informatics and the transformation of health care*. Cambridge University Press.
- Hunt, D., Haynes, B., Hanna, S., & Smith, K. (1998). Effects of Computer-Based Clinical Decision Support Systems on Physician Performance and Patient Outcomes: A Systematic Review. *Journal of the American Medical Association*, 280(15), 1339-1346.
- Kaplan, B. (2001). Evaluating informatics applications - clinical decision support systems literature review. *International Journal of Medical Informatics*, 64, 15-37.

Kappen, T. (2015). *Prediction Models and Decision Support: Chances and Challenges*. Universiteit Utrecht.

Kappen, T., Van Loon, K., Kappen, M., Van Wolfswinkel, L., Vergouwe, Y., Van Klein, W., Moons, K., Kalkman, C. (2016). Barriers and facilitators perceived by physicians when using prediction models in practice. *Journal of Clinical Epidemiology*, 70, 136-145.

Kappen, T. (2019). *Persoonlijke communicatie*. Utrecht, 5 maart.

Kawamoto, K., Houlihan, C., Balas, A., & Lobach, D. (2005). Information in practice: Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *British Medical Journal*, 1-8.
<https://doi.org/10.1136/bmj.38398.500764.8F>

Kirwan, J., Chaput de Saintonge, M., & Joyce, C. (1990). Clinical Judgment Analysis. *Quarterly Journal of Medicine*, 76(281), 935-949.

Légaré, F., Ratté, S., Gravel, K., & Graham, I. (2008). Barriers and facilitators to implementing shared decision-making in clinical practice: Update of a systematic review of health professionals' perceptions. *Patient Education and Counseling*, 73, 526-535.

Li, A.C., Kannry, J.L., Kushniruk, A., Chrimes, D., McGinn, T.G., Edonyabo, D., & Mann, D.M. (2012). Integrating usability testing and think-aloud protocol analysis with “near-live” clinical simulations in evaluating clinical decision support. *International Journal of Medical Informatics*, 81, 761-772.

Lundgrèn-Laine, H., & Salanterä, S. (2010). Think-Aloud Technique and Protocol Analysis in Clinical Decision-Making Research. *Qualitative Health Research*, 20(4), 565-575.

Luther, V.P., & Crandall, S.J. (2011). Commentary: Ambiguity and Uncertainty: Neglected Elements of Medical Education Curricula? *Academic Medicine*, 86(7), 799-800.

Macnish, K. (2012). Unblinking eyes: The ethics of automating surveillance. *Ethics and Information Technology*, 14(2), 151-167.

March, J.G. (1987). Ambiguity and Accounting: The Elusive Link Between Information and Decision Making. *Accounting, Organizations and Society*, 12(2), 153-168.

March, J. G. (2010). *The Ambiguities of Experience*. New York: Cornell University Press.

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, 6(3), 175-183.

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 1-21.

- Murdoch, T.B., & Detsky, A.S. (2013). The Inevitable Application of Big Data to Health Care. *Journal of the American Medical Association*, 309(13), 1351-1352.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.
- Nisbett, R.E., & Wilson, T.D. (1977). Telling More Than We Can Know: Verbal Reports on Mental Processes. *Psychological Review*, 84(3), 231-259.
- Noordegraaf, M., & Abma, T. (2003). Management by Measurement? Public Management Practices amidst Ambiguity. *Public Administration*, 81(4), 853-871.
- Peterson, M. (2009). *An Introduction to Decision Theory*. Cambridge University Press.
- Schön, D. (1991). *The Reflective Practitioner: How Professionals Think in Action*. Londen: Basic Books.
- Schwartz, S. & Griffin, T. (1986). *Medical Thinking: The Psychology of Medical Judgment and Decision Making*. Springer-Verlag.
- Sim, I. (2016). Two Ways of Knowing: Big Data and Evidence-Based Medicine. *Annals of Internal Medicine* (164), 562-563.
- Sim, I., Gorman, P., Greenes, R., Haynes, B., Kaplan, B., Lehmann, H., & Tang, P. (2001). Clinical Decision Support Systems for the Practice of Evidence-based Medicine. *Journal of the American Medical Informatics Association*, 8(6), 527-534.
- Sittig, D., Wright, A., Osheroff, J., Middleton, B., Teich, J., Ash, J., ... Bates, D. (2008). Grand challenges in clinical decision support. *Journal of Biomedical Informatics*, 41, 387-392.
- Steyerberg, E.W. (2009). *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. New York: Springer Science+Business Media.
- Todd, P. & Benbasat, I. (1987). Process Tracing Methods in Decision Support Systems Research: Exploring the Black Box. *Management Information Systems Quarterly*, 493-512.
- Van Bodengraven, P. (2018, maart). Evidence-based werken en de noodzaak om breder te kijken. Geraadpleegd van <https://www.waardigheidentrots.nl/verslagen/evidencebased-werken-en-de-noodzaak-om-breder-te-kijken/>
- Van der Wal, Z., De Graaf, G., & Lasthuizen, K. (2008). What's Valued Most? Similarities and Differences Between the Organizational Values of the Public and Private Sector. *Public Administration*, 86(2), 465-482.
- Van Os, L. (2009). Methodische reflectie als vorm van professionalisering. *Management & Consultancy*, 100-112.
- Van Oorschot, J. A., Jaspers, C. A., Schaaf, J. H., Linnebank, F., Oostveen, C. A. G., & Braaksma, J. T. (1995). *Professionele Autonomie van de Medisch Specialist*. Assen: Van Gorcum

Wang, Y. & Ruhe, G. (2007). The Cognitive Process of Decision Making. *International Journal of Cognitive Informatics and Natural Intelligence*, 1(2), 73-85.

Wilensky, H. (1964). The Professionalization of Everyone? *The American Journal of Sociology*, 70(2), 137-158.

Bijlagen

Bijlage A: Verkenning van de klinische praktijk – IC Wilhelmina Kinderziekenhuis Utrecht

Als vooronderzoek is er door de onderzoeker op 7 mei 2019 meegelopen op de Intensive Care van het Wilhelmina Kinderziekenhuis in Utrecht. Hier werd één van de artsen gevolgd in de ochtend van 08.00 uur tot 13.00 uur. Er is kennis gemaakt met de dagelijkse routine en de beslissingsmomenten waar de arts mee te maken krijgt.

Om 08.15 uur wordt eerst de overdracht gedaan van de nachtdienst. Hierbij wordt een update gegeven van wat er in de nacht is gebeurd, welke beslissingen er genomen zijn en in sommige gevallen welk beleid zij in gedachte hebben voor de patiënt. Vervolgens is er een ronde langs de patiënten waarbij er wordt bekeken hoe het op dat moment met de patiënten gaat, ook samen met andere behandelende artsen buiten de IC.

Om 09.00 uur is er de ‘dagstart’ met de hele afdeling waarin in een kwartier de ‘successen en verbazingen’ worden besproken. Dit gaat breder dan de patiënten en kan ook gaan over bijvoorbeeld huishoudelijke afspraken. Hierbij wordt aangegeven wat het plan van aanpak wordt en wie het gaat oppakken. Ook wordt de voortgang van eerdere acties besproken.

Vervolgens is er meegelopen met een van de behandelaars van een patiënt die een echo ging maken. Diegene heeft mij ook wat meer kunnen vertellen hoe sommige beslissingen worden genomen. Sommige beslissingen kunnen individueel worden gemaakt, bijvoorbeeld wanneer een patiënt stabiel is of iemand veel ervaring heeft met de behandeling. Er zijn veel momenten waarop beslissingen en vooral het beleid met betrekking tot een patiënt gezamenlijk worden gemaakt. Daarnaast is data of informatie niet altijd eenduidig: er kunnen bijvoorbeeld verschillen zitten in metingen. Zo werd er het voorbeeld gegeven van een echo waarbij de ene behandelaar een echo van de oksel tot de buik maakt, maar een ander alleen een echo maakt van de buik. Dit heeft gevolgen voor hoe het beeld eruitziet en wat de vervolgbehandeling wordt. Dit werd geconstateerd doordat de vervolgbehandeling ineens een andere richting kreeg en er kon worden teruggekeken naar de echo's die gemaakt waren. Met mondeling navragen kwamen ze erachter dat er een andere methode werd gehanteerd in het maken van de echo's.

Hierna volgde een ronde op de afdeling waarbij werd overlegd wat het beleid voor die dag ging worden voor de patiënten. Dit werd gedaan aan de hand van iemand die het verhaal van de patiënt en het verloop van de gezondheid samenvat en iemand die het elektronisch patiëntendossier (EPD) daarin aanpast en bijhoudt. Daarin wordt overlegd wat het beleid wordt waarbij aanpassingen van medicatie direct worden doorgevoerd in het EPD. Beslissingen hierin worden zowel gebaseerd op ervaringen met andere patiënten, hoe de patiënt eruitziet of erbij ligt en wat het verloop van de gezondheid is geweest en individuele kenmerken (bijvoorbeeld leeftijd) als op medische kennis. Veel informatie komt ook uit handgeschreven notities, naast de waardes van bijvoorbeeld bloeddruk of lab-uitslagen die in het EPD staan weergegeven.

Na deze ronde is er meegelopen met een van de IC'ers naar de afdeling radiologie om enkele foto's te bekijken van patiënten. Hierbij deed de radioloog uitspraak over de situatie en in overleg met de aanwezige behandelaars van de patiënt werd oorzaken en mogelijk beleid besproken. Daarna volgde een bespreking met alle zorgverleners van de IC en behandelaars van andere afdelingen (zoals cardiologie) over de patiënten. Daarin werd met iedereen besproken hoe het gaat met de patiënten en wat het beste vervolgbeleid is. Hierin lag met name de nadruk op de input van de behandelaars van andere afdelingen.

Deze bespreking werd vervolgd met een andere bespreking over een aantal specifieke patiënten, waarvan met name één de aandacht had omdat het een complexe casus was. Hierin waren veel zorgverleners van zowel de IC als andere afdelingen aanwezig en werd de hele situatie van de patiënt besproken, zowel medisch als sociaal. Overwegingen in de behandeling werden besproken en bediscussieerd om tot de beste oplossing te komen. Hierin speelt niet alleen de medische context een rol (wat is de beste behandeling voor de patiënt, waarin ieder vanuit zijn of haar specialisme beredeneert wat de beste behandeling zou kunnen zijn) maar ook de sociale context (wat is haalbaar voor deze patiënt gezien de sociale situatie). Ook ethische overwegingen kwamen aan bod.

Mijn bevindingen betreffende de klinische praktijk op de IC is dat veel beslissingen in overleg worden gemaakt. Er is naast 'rationele besluitvorming' (uit de metingen blijkt deze ziekte en dat vraagt om behandeling X vanuit medische kennis) ook sprake van veel klinische besluitvorming: er wordt gekeken naar hoe de patiënt erbij ligt en hoe dat ten opzichte van voorgaande dagen is, informatie vanuit ouders wordt meegenomen (een druk bewegend kind is minder afwijkend wanneer het thuis ook al druk is) en ervaring met soortgelijke patiënten uit het verleden wordt meegewogen in de behandeling. In sommige gevallen wordt ook de sociale omgeving meegewogen voor het besluit over de vervolgbehandeling. Er zijn grofweg twee soorten data te onderscheiden: waardes vanuit metingen (lab-uitslagen, bloeddrukmetingen, etc.) en getypte of handgeschreven observaties en bevindingen van de behandelaars. Beide worden gebruikt om (gezamenlijk) tot de beste beslissing te komen voor een patiënt waarin de behandelaars middels argumentatie met elkaar tot een overeenstemming komen.

Bijlage B: Topiclijst interviews ontwikkelaars

Opening

Bedankt voor jullie tijd voor dit interview. Ik ben student Bestuurs- en Organisationswetenschap en ik doe onderzoek naar hoe medische data betekenis krijgt met betrekking tot bias in de app UrSmartStatus. Hierbij kijk ik zowel naar hoe er wordt omgegaan met bias en hier betekenis aan wordt gegeven door de ontwikkelaars als naar de interpretatie daarvan van de toekomstige gebruikers, artsen. De bias met betrekking tot ambiguïteit en onzekerheid van medische data.

Het interview zal opgenomen worden en getranscribeerd, waarna de opname verwijderd zal worden. Het interview is volledig geanonimiseerd. Geven jullie hier toestemming voor?

Voorstellen respondenten

- Hoe lang zijn jullie al betrokken bij het project?
- Hoe zouden jullie omschrijven wat jullie doen en wat er al is gedaan?
- Lopen jullie tegen dingen aan (mbt bias/besluitvorming)?

Bias

- In het evaluatierapport staat: “Het faciliteert de zorgverleners om snel en efficiënt de status van patiënten vast te stellen en daarbij meer inzicht te hebben in mogelijk ontbrekende informatie (bias)”. Hoe wordt de bias in kaart gebracht/inzichtelijk gemaakt in de app voor de gebruikers?
- Hoe ga je om met ambiguïteit van medische data? Hoe komt dit tot uiting in de app?
- Hoe wordt dit weergegeven in de app? Waarom?
- Op welke manier kan de app weten wanneer er gebrek aan informatie is?
- Op welke manier is dit inzichtelijk voor de arts, wanneer er gebrek aan informatie is? Of wordt dit overgelaten aan het inzicht van de arts?
- Hoe geef je weer als er nog geen diagnose kan worden gegeven?
- Hoe ga je om met onzekerheid van medische data, gebrek aan voorspelbaarheid?
- Waarop is een prognose gebaseerd? Waarom?
- Hoe bepaalt het algoritme de prognose? (historische patiëntgevallen, individuele medische data?)
 - o Als prognose op basis van historische patiëntgevallen is, wanneer is dit dan een gedegen voorspeller van de gezondheid van de huidige patiënt en hoe moet de arts dit bepalen? Hoe voorkom je dat dit leidraad wordt? Wordt dit overgelaten aan de arts?
- Hoe geef je weer als er geen prognose kan worden gegeven door bv. gebrek aan data?
- In de evaluatie is een boom gemaakt met waaruit bias van data in de zorg bestaat. Hoe verhouden ambiguïteit en onzekerheid hiertoe?
- Hoe gaan jullie met al deze problemen om? Hoe uit zich dat in de app?
- Waarom zijn inzicht in de data en ontbrekende data gescheiden? Hebben deze niet met elkaar te maken?
- Wat wordt bedoeld met externe invloeden en artefacten in de data?
- Hoe gaat de app om met ‘fouten’ van invoering van data (verkeerd geschreven o.i.d.), wordt dit inzichtelijk voor de arts?
- Bij ontbrekende data: hoe kan het dat data niet benaderbaar is? (als het gemeten is, is het toch ook ergens opgeslagen?)

Algoritme versus medicus (besluitvorming)

- Hoe worden keuzes gemaakt in de app, op basis waarvan? Overleg, wetenschappelijke literatuur, is er verantwoording op de keuzes? Heb je daar een voorbeeld van, hoe iets tot stand is gekomen? Waar is lang overleg over geweest, waar kwamen jullie niet uit?
- Hoe zou je het verschil omschrijven tussen het gebruik van het EPD en de app?
- Hoe zou je het verschil uitleggen tussen de interpretatie van data door de app/algoritme en de interpretatie door een arts? (want neemt een deel over van wat de arts normaal gesproken doet) Hoe gaan jullie hiermee om in het ontwikkelen van de app?
- Verandert het de besluitvorming van de arts? Op welke manier?
- Hoe is de verwerking/analyse van het algoritme op de data 'objectief' (uit evaluatie)?
- Hoe werkt de rationele besluitvorming van het algoritme?
- Hoe sluit dit aan op de besluitvorming van artsen, die naast data ook veel klinische (subjectieve) besluitvorming doen?
- Op welke manier is er aandacht voor hoe artsen data verwerken, bv. dat statistische data lastig te verwerken is?

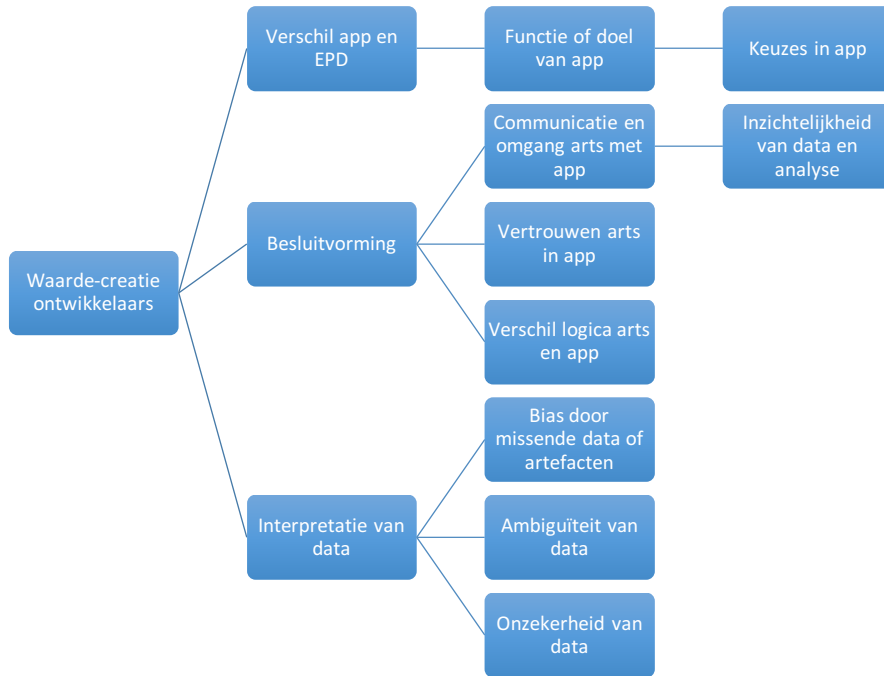
Toelichting/validatie

- Geeft het model toelichting op de status van de patiënt door toelichting te geven op waar de status door bepaald is? Hoe? Hoe ziet dat eruit? Data + analyse? Vind je nodig dat er toelichting wordt gegeven?
- Als het doel is om de arts snel te informeren, moet hij toch ook vertrouwen op hoe de informatie aangeboden wordt, en zou hij in principe niet naar de toelichting kijken? Hoe verhoudt zich dit tot elkaar?
- Uit het evaluatierapport: "beter uitlegbare ernst-interpretatie, het model is beter uitlegbaar geworden" → hoe?

Afsluiting

- Hoe zouden jullie omschrijven hoe er betekenis wordt gegeven aan ambiguïteit en onzekerheid van medische data in de app, of hoe dit in de toekomst zou moeten gebeuren?
- Willen jullie zelf nog iets toevoegen of is er iets waar ik niet naar gevraagd heb, maar wat in jullie ogen wel belangrijk is?
- Bedankt voor jullie medewerking, tijd en informatie. Mochten er na dit interview nog vragen of opmerkingen zijn, kunnen jullie mij mailen.

Bijlage C: Codeboom interviews ontwikkelaars



Bijlage D: Uitleg voor de arts respondent in experiment

Fijn dat u wilt deelnemen aan dit onderzoek. Het onderzoek is in uitvoering van de Universiteit Utrecht en het doel is om het besluitvormingsproces van artsen in verschillende situaties in kaart te brengen. Het onderzoek zal ongeveer 45 minuten duren. Ik zal straks 2 verschillende patiëntcasussen voorleggen waarbij de gegevens van de patiënt de ene keer te vinden zijn in een app en de andere keer in HiX en Metavision. Voordat we beginnen zal ik u ook de app en haar functies tonen. Hierbij zal ik u vragen om uw denkproces hardop uit te spreken, dus elke stap die u in uw hoofd zet om tot een besluit/diagnose/beleid te komen. Dat kunnen overwegingen zijn, besluiten om bepaalde informatie te bekijken, vragen die er bij u opkomen, etc. Ik zal hier straks ook een voorbeeld van geven hoe zo een hardop uitgesproken denkproces kan gaan.

Het is de bedoeling dat u zich inbeeldt dat het gaat om patiënten die al een aantal dagen bij u op de afdeling liggen. Het is een overdrachtsmoment waarbij de voorgeschiedenis en informatie over de patiënt gegeven wordt in een samenvatting. Aan de hand van die informatie kan u de gegevens van de patiënt ofwel in HiX of in de app bekijken en hardop uitspreken wat u denkt. Net zoals in de werkelijkheid gaat u de patiënt diagnosticeren en besluiten welk beleid en vervolgstappen nodig zijn.

Dit is hoe de app eruitziet. De app geeft een visuele weergave en algemeen beeld van de gezondheid van de patiënt aan de hand van de organen. Deze kleuren rood, oranje, geel of groen. Bij groen is het orgaan min of meer stabiel en geel, oranje en rood geven een gradatie van verslechtering aan. Lichtgroen en lichtoranje betekenen dat er mogelijk een gebrek aan informatie is, door bijvoorbeeld oude waarden of te weinig informatie. U kan op de organen klikken om meer informatie te verkrijgen. Hier ziet u hoe het met het orgaan gaat in een tijdsverloop van opname naar nu waarbij ook de bijbehorende parameters te zien zijn. De kleuren zijn gebaseerd op de patiëntgegevens en de data analytics die de urgentie samenvat.

De samenvatting wordt hier gegeven, en met de knop 'status' kunt u doorklikken naar HiX en Metavision. Daarin is een omgeving gemaakt wat een combinatie is tussen HiX en Metavision, zoals je hier in de tabbladen kunt zien. Dit zijn de verschillende tabbladen. In het journaal kun je scrollen door op de pijltjes boven en onder te klikken, je kan de balk niet verschuiven. De laatste conclusie is weggelaten voor deze casus. Hier is ook de samenvatting te vinden. Met de laatste knop kunt u terug naar de app.

Het hardop uitspreken van het denkproces zou bijvoorbeeld op de volgende manier kunnen gaan. Ik krijg de volgende informatie: *het gaat om een 18-jarige jongen die met acute buikpijn en verhoogde infectiewaardes is opgenomen en op dag 1 is geopereerd voor een blindedarmonsteking. De afgelopen dagen ging het goed en hij zou vandaag ontslag krijgen, maar nu krijgt hij koorts en voelt hij zich niet lekker.* Ik ga kijken naar de hartfrequentie, ik zie dat die oploopt. Dan bekijk ik de ademhalingsfrequentie, die loopt ook op. Even kijken hoe hoog zijn temperatuur is, die is 40.2 dus dat is inderdaad verhoogd. Ik vraag me af of er al een echo is gemaakt, eens zien. Nee die is nog niet gemaakt dus dat zou nog kunnen. Zijn er nog nieuwe waarden uit het lab, ook niet. Waar ik in eerste instantie aan denk is een abces bij de appendix en daar wil ik een echo voor laten maken en bloed afnemen.

Als u straks een casus in de app krijgt aangeboden, wil ik u vragen om in ieder geval in eerste instantie de app te gebruiken om tot een eerste indruk, conclusie of beleid te komen. U mag ten alle tijden het EPD erbij halen om meer informatie te verkrijgen. De app is ook geen vervanging van het EPD, u kan

ze naast elkaar gebruiken. HiX mag altijd gebruikt worden! De datum die er staat mag je inbeelden dat dat vandaag is.

Het experiment zal opgenomen worden met schermopname en audio. Het kan zijn dat ik straks tussendoor af en toe vraag om hardop te blijven nadenken als het even stil valt of ik niet weet wat er gebeurt. Ook kan het zijn dat ik aantekeningen maak. Daarnaast heb ik nog wat vragen achteraf. Was dit duidelijk of heeft u nog vragen?

Bijlage E: Aandachtspunten tijdens het experiment en follow-up vragen

Aandachtspunten tijdens het experiment

- (On)duidelijkheden in de app
- Wordt ambiguïteit van data wel of niet duidelijk in de app
- Staat de respondent wel of niet open voor meerdere interpretaties
- Zijn er (on)duidelijkheden over functies van de app
- Mate van vertrouwen in de app
- In hoeverre wordt er DD gegeven/conclusie getrokken bij app, hoe snel wordt er overschakelt naar EPD, hoeveel meer informatie is er uit het EPD nodig
- Uitdrukking van de arts: bijvoorbeeld bedenkelijk/vragend kijken

Bij stiltes wordt er gevraagd ‘Wat denkt u nu?/Wat bent u nu aan het doen?/U kunt hardop blijven nadenken’. Bij het uitblijven van verheldering over wat er wordt gedaan (respondent wisselt van tabblad maar zegt dit niet) wordt ook gevraagd ‘Wat gaat u nu doen?/Wat heeft u nu gedaan?’

Na elke casus de vraag:

- Wat is uw differentiaal diagnose over deze patiënt en wat zijn de vervolgstappen?

Follow-up vragen

- Hoe staat u tegenover het gebruik van een dergelijke decision-support app?
- Wat vond u van het gebruik? Wat zijn uw positieve ervaringen, wat zijn uw negatieve ervaringen?
- Vond u de app gebruiksvriendelijk?
 - o Complex, gemakkelijk, inconsistenties, lastig, zelfverzekerd in gebruik, meer info vooraf nodig
 - o Waren de functies voor de hand liggend, logisch?
- Had u vertrouwen in het systeem dat de medische data goed geïnterpreteerd/verwerkt was?
- Had u het idee dat u de rood en oranje gekleurde organen bekeek door de kleuren, dat u zonder de app wellicht niet naar deze organen had gekeken?
- Zou u de waardes en/of uitkomsten van de app checken door in het EPD te kijken? Standaard of bij uitzonderlijke gevallen?
- Zou u de app als toegevoegde waarde zien naast HiX/Metavision, waarom wel/niet?
- Als de app in de dagelijkse praktijk zou werken, zou u hem dan gebruiken? Wanneer wel/niet?
- Demografische gegevens
 - o Functie
 - o Aantal jaren werkend (in deze functie)
 - o Geslacht
 - o Leeftijd

Bijlage F: Patiëntcasus 1

Voor het tonen van casus 1 zijn enkele afbeeldingen gemaakt van de tabbladen die de arts te zien kreeg in zowel de applicatie als in het EPD. Beschrijving van de casus:

Ernst van de Leerkamp, 15 weken oude zuigeling, bekend met ccTGA met DORV, inlet VSD en hypoplastische boog wv boogrepair en banding van de arterie pulmonalis op 18/2/2019. Herstellende van hartfalen, waarvoor tot een week terug opgenomen op de PICU.

Situatie: Nu in nachtdienst naar afdeling Leeuw geroepen vanwege respiratoire achteruitgang

The screenshot shows the 'smartstatus' application interface. At the top left, there is a 'smartstatus' logo and a '< Terug' button. The main content area is divided into two columns. The left column contains patient information: 'Ernst van de Leerkamp', 'PATIENT-ID 1234567', 'GEWICHT 5650 gram', 'GEBORTE DATUM 12-02-2019', and 'LAATSTE UPDATE 18-05-2019 06:21'. Below this is a detailed medical history paragraph and a 'STATUS' button. The right column features a blue header with navigation tabs (-7D, -3D, -24U, -8U, 0U) and a central graphic of a house with icons representing different medical systems (heart, lungs, stomach, etc.).

This screenshot shows the 'smartstatus' application with the 'RESPIRATIE' (Respiration) tab selected. The patient information on the left is identical to the previous screenshot. The central graphic is now titled 'RESPIRATIE' and highlights the respiratory system icons. The right column displays a list of respiratory parameters: 'SpO2 89%', 'Adem.frequentie 65 x/min toegenomen', 'Intrekkingsen subcostaal', 'Luchtweg snotterig; hoesten', 'FI02 21%', 'pH (cap) 7.42', 'pO2 (cap) 42 mmHg', 'pCO2 (cap) 40 mmHg', and 'X-thorax Verricht: zie HIX'. A '< Terug' button is visible at the top left.

< Terug

Ernst van de Leerkamp

PATIENT-ID 1234567 GEWICHT 5650 gram

GEBOORTE DATUM 12-02-2019

LAATSTE UPDATE 18-05-2019 06:21

15 weken oude zuigeling, bekend met cTGA met DORV, inlet VSD en hypoplastische boog wv boogrepar en banding van de arterie pulmonalis op 18/2/2019. Herstellende van hartfalen, waarvoor tot een week terug opgenomen op de PICU.

Nu in nachtdienst naar afdeling Leeuw geroepen vanwege respiratoire achteruitgang

STATUS

-7D -3D -24U -8U 0U

CIRCULATIE

CIRCULATIE

-7D -3D -24U -8U 0U

Bloeddruk -

Hartfrequentie 153 x/min

SpO2 93%

Lactaat -

BNP 41 pmol/L

Kleur kind bleek

Furosemide 2 mg/kg/dag

Vochtbalans +340 ml

< Terug

Ernst van de Leerkamp

PATIENT-ID 1234567 GEWICHT 5650 gram

GEBOORTE DATUM 12-02-2019

LAATSTE UPDATE 18-05-2019 06:21

15 weken oude zuigeling, bekend met cTGA met DORV, inlet VSD en hypoplastische boog wv boogrepar en banding van de arterie pulmonalis op 18/2/2019. Herstellende van hartfalen, waarvoor tot een week terug opgenomen op de PICU.

Nu in nachtdienst naar afdeling Leeuw geroepen vanwege respiratoire achteruitgang

STATUS

-7D -3D -24U -8U 0U

GASTRO-INTESTINAAL

GASTRO-INTESTINAAL

-7D -3D -24U -8U 0U

Voeding (NMS) 20 ml/h

Voeding (DDS) 5 ml/h

Retenties mondje

Glucose 5.8 mmol/L

Gewicht 5650 gram

Colecalciferol 400 IE/dag

Kapfen, T.M. - Productie omgeving - 6.1.1173 - FAX - ChpSet

Kinder algemene pedatrie
05-02-2019

Leerkamp, Ernst van de
12-02-2019 (16 - 07 wk)
1265497
061112322

Behandelaar
Kinder algemene pedatrie
Klinisch hoofdbehandelaar
Kinderendocrinologie

Risico vankomst
Overname Pellicaan

Kerngegevens
Contactpersonen
Type: Jansen, K.J. - Moder van
Tele: 061112322
Type: Leerkamp, D.K. de - Vader van de
Tele: 062221111
Portaal gemachtigden
K.J. Jansen, V. 12-06-1981
D.K. van de Leerkamp, M. 14-05-1980

Allergieën en bijwerkingen
PLEESTERS/ALEFLAAG
Med
Medicatie
COLECALCFEROLVITAMINE D3 40E = 1 MICROG/DRUPEL (ORAAL, 1 x per dag 400 internat.eenh.)

Complicaties per specialisme (laatste 5)

Aandachtspunten
Keuzelijst - 21-05-2019: Kliniek: vervolgonderzoek

Behandeldoelen

Samenvatting
Dossier samenvatting (18-05-2019)
15 weken oude zuigeling, bekend met cTGA met DORV, iniet VSD en hypoplastische boog in booggebeel en banding van de atere pulmonals op 18/02/2019. Herstellende van hartfalen, waanvoor tot een week terug opgenomen op de PICU.
18-05-2019 06:21 uur:
Nu in revalidatie naar afdeling Leneu groepen vanwege respiratoire achteruitgang

Anamnese
17-05-2019 11:39 KVC
Korperhoe...
74616 Korperhoeak
Patiënt in de ochtend beoordeeld samen met W. van Dijk. Tachydyssrhythmisch met AHF rond 50/min en head bobbing. Periferie goed gecirculeerd. Maakt hierbij een ontspannen indruk, ligt helder om zich heen te kijken.
17-05-2019 08:54 KVC
Dirksen, E...
KAP SEMMA
Hoofdbehandelaar: KCA
Semi-arts: E. Dirksen 75077
Dagsupervisor: Korperhoeak
Reden van opname: status na booggebeel en banding MPA op 18-2 bij cTGA met DORV
Ontslagcriteria: hemodyn. ernstig stabiel, voeding volledig over de maag
Vitalie
Resp: tachydyssrhythmisch met AH 60-70, subcostale intrekkingen, saturaties 92%
Circ: HA 165-185, in rust lager, transparant en is bleek, G -100g
GL: voeding goed verdragen, soms een mondje spugen
Int: 137.5
Neur: slecht geslapen, erg onrustig
Overig:
16-05-2019 12:17 KVC
Ijeren, G...
KCA ARTS
Anamnese
Vericht onderzoek: echo cardiografie, Echo cardiografie: Vanuit eerder bekende cTGA met DORV + PA banding Goede biventriculaire contractiliteit. Kleinere links gelegen coronair. Geringe Woot matige TI.
Grotere links gelegen moef LV. Geringe MI.
ActV ligt anterior en links van PV.
is kleiner dan PV. (Kort uit li gelegen kleinere RV, maar dit op deze echo niet aangetoond)
ActV. Vmax: 1 m/s. Ao boog nu niet in beeld.
PA komt centraal uit RV (overstijgend volledig uit RV). Voorwaartse flow vanuit LV naar PA (door VSD).
PA: turbulente flow, Vmax: > 4 m/s, over de banding. Turbulentie begint al bij de PV.
ASD met li te shunt. VSD flow vanuit LV door VSD naar Pulm.art.
AS / Act

Anamnese
16-05-2019 07:56 KVC
Van Dijk
KAP ARTS
Hoofdbehandelaar: KCA
Zaakarts: van Dijk 75078
Dagsupervisor: van Ijeren
Reden van opname: status na booggebeel en banding MPA op 18-2 bij cTGA met DORV
Ontslagcriteria: hemodynamisch stabiel, voeding volledig over de maag
Vitalie
Alp: gaat redelijk, prestatie evenwichtig
Resp: wankel evenwichtig, met vagen forse tachypnoe met veel ademarbeid, saturaties stabiel rond 90%. Geen saturatiepiek bij onrust. Kliniek vol. Hoer van snot en slijm.
Circ: transparant veel, G = 15 g, plaat goed
GL: verdraagt voeding goed over de maag, ontlasting op gang

Kapfen, T.M. - Productie omgeving - 6.1.1173 - FAX - ChpSet

Kinder algemene pedatrie
05-02-2019

Leerkamp, Ernst van de
12-02-2019 (16 - 07 wk)
1265497
061112322

Behandelaar
Kinder algemene pedatrie
Klinisch hoofdbehandelaar
Kinderendocrinologie

Risico vankomst
Overname Pellicaan

Kerngegevens
Contactpersonen
Type: Jansen, K.J. - Moder van
Tele: 061112322
Type: Leerkamp, D.K. de - Vader van de
Tele: 062221111
Portaal gemachtigden
K.J. Jansen, V. 12-06-1981
D.K. van de Leerkamp, M. 14-05-1980

Allergieën en bijwerkingen
PLEESTERS/ALEFLAAG
Med
Medicatie
COLECALCFEROLVITAMINE D3 40E = 1 MICROG/DRUPEL (ORAAL, 1 x per dag 400 internat.eenh.)

Complicaties per specialisme (laatste 5)

Aandachtspunten
Keuzelijst - 21-05-2019: Kliniek: vervolgonderzoek

Behandeldoelen

Samenvatting
Dossier samenvatting (18-05-2019)
15 weken oude zuigeling, bekend met cTGA met DORV, iniet VSD en hypoplastische boog in booggebeel en banding van de atere pulmonals op 18/02/2019. Herstellende van hartfalen, waanvoor tot een week terug opgenomen op de PICU.
18-05-2019 06:21 uur:
Nu in revalidatie naar afdeling Leneu groepen vanwege respiratoire achteruitgang

Anamnese
17-05-2019 11:39 KVC
Korperhoe...
74616 Korperhoeak
Patiënt in de ochtend beoordeeld samen met W. van Dijk. Tachydyssrhythmisch met AHF rond 50/min en head bobbing. Periferie goed gecirculeerd. Maakt hierbij een ontspannen indruk, ligt helder om zich heen te kijken.
17-05-2019 08:54 KVC
Dirksen, E...
KAP SEMMA
Hoofdbehandelaar: KCA
Semi-arts: E. Dirksen 75077
Dagsupervisor: Korperhoeak
Reden van opname: status na booggebeel en banding MPA op 18-2 bij cTGA met DORV
Ontslagcriteria: hemodyn. ernstig stabiel, voeding volledig over de maag
Vitalie
Resp: tachydyssrhythmisch met AH 60-70, subcostale intrekkingen, saturaties 92%
Circ: HA 165-185, in rust lager, transparant en is bleek, G -100g
GL: voeding goed verdragen, soms een mondje spugen
Int: 137.5
Neur: slecht geslapen, erg onrustig
Overig:
16-05-2019 12:17 KVC
Ijeren, G...
KCA ARTS
Anamnese
Vericht onderzoek: echo cardiografie, Echo cardiografie: Vanuit eerder bekende cTGA met DORV + PA banding Goede biventriculaire contractiliteit. Kleinere links gelegen coronair. Geringe Woot matige TI.
Grotere links gelegen moef LV. Geringe MI.
ActV ligt anterior en links van PV.
is kleiner dan PV. (Kort uit li gelegen kleinere RV, maar dit op deze echo niet aangetoond)
ActV. Vmax: 1 m/s. Ao boog nu niet in beeld.
PA komt centraal uit RV (overstijgend volledig uit RV). Voorwaartse flow vanuit LV naar PA (door VSD).
PA: turbulente flow, Vmax: > 4 m/s, over de banding. Turbulentie begint al bij de PV.
ASD met li te shunt. VSD flow vanuit LV door VSD naar Pulm.art.
AS / Act

Anamnese
16-05-2019 07:56 KVC
Van Dijk
KAP ARTS
Hoofdbehandelaar: KCA
Zaakarts: van Dijk 75078
Dagsupervisor: van Ijeren
Reden van opname: status na booggebeel en banding MPA op 18-2 bij cTGA met DORV
Ontslagcriteria: hemodynamisch stabiel, voeding volledig over de maag
Vitalie
Alp: gaat redelijk, prestatie evenwichtig
Resp: wankel evenwichtig, met vagen forse tachypnoe met veel ademarbeid, saturaties stabiel rond 90%. Geen saturatiepiek bij onrust. Kliniek vol. Hoer van snot en slijm.
Circ: transparant veel, G = 15 g, plaat goed
GL: verdraagt voeding goed over de maag, ontlasting op gang

Test

Test	01-05-2019	01-05-2019	02-05-2019	03-05-2019	06-05-2019	13-05-2019	13-05-2019	16-05-2019	18-05-2019	18-05-2019	Ref. waarde	Eenh.
	05:44	06:50	06:50	11:40	08:00	08:00	09:17	09:09	09:00	06:19		
	PKCKCC	PKCKCC	PKCKCC	PKCKCC	PKCKCC	AMCA	AMCA	AMCA	AMCA	AMCA		
Opmerkingen												
Bloedgasen												
Opmerking order	Venue											
Materiaaltype					Capillair	Capillair	Capillair					
Zuurgraad (Ar)	7.39				<Memo> 7.40	7.41	7.42				7.37 - 7.45	
Foodzuurspanning (Ar)	42				<Memo> 48	41	40				35 - 45	mm Hg
Zuurloospanning (Ar)	39				<Memo> 39	41	42				20 - 100	mm Hg
Actueel bicarbonaat (Ar)	27.5				<Memo> 30.1	28.0	25.7				22.0 - 29.0	mmol/L
Bise Electro (Ar)	2.5				<Memo> 5.0	3.8	1.0				-3.0 - 3.0	mmol/L
Standaard bicarbonaat (Ar)	25.1				<Memo>							mmol/L
Zuurstof Saturatie (Ar)	0.48				<Memo>							mmol/mol
Zuurgraad (Ven)		7.42										mm Hg
Foodzuurspanning (Ven)		42										mm Hg
Zuurloospanning (Ven)		37										mm Hg
Actueel bicarbonaat (Ven)		27.2										mmol/L
Bise Electro (Ven)		2.8										mmol/L
Standaard bicarbonaat (Ven)		25.6										mmol/L
Zuurstof Saturatie (Ven)		0.49										mmol/mol
Biochemie												
Natrium						135	135				137	mmol/L
Kalium						5.1	4.9				4.5	mmol/L
Calcium geïoniseerd						1.30					1.05 - 1.50	mmol/L
Ureum						3.2	3.1				3.0 - 7.5	mmol/L
Creatinine						0.7	0.8				0.2 - 0.8	mmol/L
Lactaat											0.0 - 2.2	mmol/L
CRP											0 - 10	mg/L
BNP		213	H	215	H	25	27	H	41	14	28	pmol/L
Glucose												
Glucose							4.9			5.8		mmol/L
Geneesmiddelen												
Vancomycine		20.6										mg/L
Vancomycine (DS)		<Memo>										

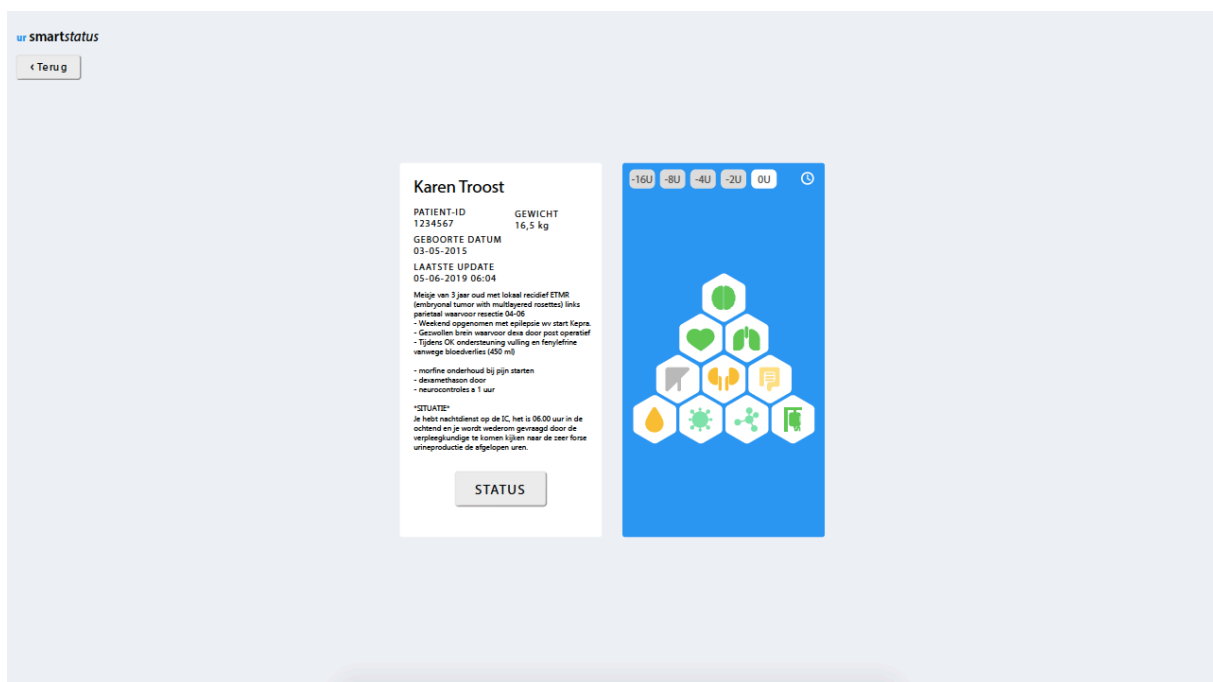
Bijlage G: Patiëntcasus 2

Voor het tonen van casus 2 zijn enkele afbeeldingen gemaakt van de tabbladen die de arts te zien kreeg in zowel de applicatie als in het EPD. Beschrijving van de casus:

Karen Troost, meisje van 3 jaar oud met lokaal recidief ETMR (embryonal tumor with multilayered rosettes) links parietaal waarvoor resectie 04-06.

- Weekend opgenomen met epilepsie wv start Kepra.
- Gezwollen brein waarvoor dexta door post operatief
- Tijdens OK ondersteuning vulling en fenylefrine vanwege bloedverlies (450 ml)
- morfine onderhoud bij pijn starten
- dexamethason door
- neurocontroles a 1 uur

Situatie: Je hebt nachtdienst op de IC, het is 06.00 uur in de ochtend en je wordt wederom gevraagd door de verpleegkundige te komen kijken naar de zeer forse urineproductie de afgelopen uren.



The image shows a screenshot of a digital interface for patient status. On the left, a white card displays patient details for Karen Troost, including her ID (1234567), weight (16.5 kg), birth date (03-05-2015), and last update (05-06-2019 06:04). It also contains a detailed medical history and a 'SITUATIE' section describing the current clinical situation. On the right, a blue mobile app interface is shown, featuring a top navigation bar with time slots (-16U, -8U, -4U, -2U, 0U) and a central grid of icons representing various medical functions like heart rate, lungs, and medication.

Karen Troost

PATIENT-ID: 1234567 GEWICHT: 16,5 kg

GEBOORTE DATUM: 03-05-2015

LAATSTE UPDATE: 05-06-2019 06:04

Meisje van 3 jaar oud met lokaal recidief ETMR (embryonal tumor with multilayered rosettes) links parietaal waarvoor resectie 04-06.

- Weekend opgenomen met epilepsie wv start Kepra.
- Gezwollen brein waarvoor dexta door post operatief
- Tijdens OK ondersteuning vulling en fenylefrine vanwege bloedverlies (450 ml)
- morfine onderhoud bij pijn starten
- dexamethason door
- neurocontroles a 1 uur

SITUATIE:
Je hebt nachtdienst op de IC, het is 06.00 uur in de ochtend en je wordt wederom gevraagd door de verpleegkundige te komen kijken naar de zeer forse urineproductie de afgelopen uren.

STATUS

< Terug

Karen Troost

PATIENT-ID: 1234567 GEWICHT: 16,5 kg

GEBORTE DATUM: 03-05-2015

LAATSTE UPDATE: 05-06-2019 06:04

Meisje van 3 jaar oud met lokaal recidief ETMR (embryonale tumor with multilayered rosettes) links parietaal waartoevoren resectie 04-06

- Weekend opgenomen met epilepsie wv start Kepra
- Gewelven brein waartoevoren dieet door post-operatief
- Tijdens OK ondersteuning vulling en fenylefrine vanwege bloedverlies (450 ml)
- monitorie onderhouden bij pijn starten
- desamethason door
- neurocontroles a 1 uur

'SITUATIE'
Je hebt nachtdienst op de IC, het is 06:00 uur in de ochtend en je wordt wederom gevraagd door de verpleegkundige te komen kijken naar de zeer forse urineproductie de afgelopen uren.

STATUS

-16U -8U -4U -2U 0U

CIRCULATIE

CIRCULATIE ❤️

OP OK - NIET IC

-16U

-8U

-4U

-2U

0U

Bloeddruk 81/41 (62) mmHg

Hartfrequentie 93 x/min

SpO2 100%

Fenylefrine 100 mcg/ml 3 ml/uur

Diurese laatste uur 45 ml

Bloedverlies 450 ml

< Terug

Karen Troost

PATIENT-ID: 1234567 GEWICHT: 16,5 kg

GEBORTE DATUM: 03-05-2015

LAATSTE UPDATE: 05-06-2019 06:04

Meisje van 3 jaar oud met lokaal recidief ETMR (embryonale tumor with multilayered rosettes) links parietaal waartoevoren resectie 04-06

- Weekend opgenomen met epilepsie wv start Kepra
- Gewelven brein waartoevoren dieet door post-operatief
- Tijdens OK ondersteuning vulling en fenylefrine vanwege bloedverlies (450 ml)
- monitorie onderhouden bij pijn starten
- desamethason door
- neurocontroles a 1 uur

'SITUATIE'
Je hebt nachtdienst op de IC, het is 06:00 uur in de ochtend en je wordt wederom gevraagd door de verpleegkundige te komen kijken naar de zeer forse urineproductie de afgelopen uren.

STATUS

-16U -8U -4U -2U 0U

NIEREN

NIEREN 👂

-16U

-8U

-4U

-2U

0U

Creatinine 28 umol/L

Ureum 2,5 mmol/L

Diurese laatste uur 25 ml

Diurese per uur 7,7 ml/kg/uur

Vochtbalans -106 ml

