

Numbers telling the tale?

On the validity of patient experience surveys and the usability of their results

Is meten weten?

Over de validiteit van patiëntervaringsvragenlijsten en de bruikbaarheid van hun resultaten

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan Tilburg University
op gezag van de rector magnificus, prof. dr. E.H.L. Aarts,
in het openbaar te verdedigen ten overstaan van een door het college voor
promoties aangewezen commissie in de aula van de Universiteit
op vrijdag 12 juni 2015 om 14.15 uur

door

Maarten Watse Krol
geboren op 14 augustus 1984
te Herwen en Aerdt

Summary

Patient experiences have become an important part of quality of care information. In the Netherlands, this information is used by three main stakeholders in a system of regulated competition: by patients to choose a healthcare provider, by healthcare providers to improve healthcare and to account for their provided care, and by health insurance companies in purchasing healthcare (Enthoven and Van de Ven, 2007). Valid, reliable and usable information on quality of care, including quality of care from the patients' perspective, was deemed central to the system by all stakeholders.

In the quest for measuring Dutch patient experiences in a more standardized way, since 2006, patient experience surveys of the Consumer Quality Index (CQ-index or CQI) have been developed in the Netherlands (Delnoij *et al.*, 2006; 2010). An overview of the studies in its first five years of existence gave a rather positive picture of the CQI, but also some points of concern and opportunities for improvement (Hopman *et al.*, 2011).

With respect to the survey methodology, issues arose concerning the validity of patient experience surveys and the usability of survey results. The validity of the surveys hinges on the extent to which the surveys include the most relevant aspects of quality of care (face and content validity) and whether the survey items measure what they intend to measure (construct validity) (Streiner and Norman, 1999c; Mokkink *et al.*, 2012). The usability of patient experience surveys entails whether stakeholders are able to use the survey results. For example, are they able to interpret the results that are presented? Can they act upon these results to choose a healthcare provider (patients), to improve care (healthcare providers) or to purchase good quality care (health insurers)?

These are important issues with regard to measuring and improving quality of care from the perspective of patients. Also, it seems that quality of care information on patients' experiences does not always meet the demands of stakeholders. This acknowledges the relevance of further research to better understand and, consecutively, improve the validity of patient experience surveys and the usability of their results. In this thesis, six studies were presented on these issues. This thesis sought to answer two general questions:

1. *How can the validity of patient experience surveys be improved?*
2. *How can the usability of patient survey results be improved for stakeholders?*

Chapter 2 focused on the content validity of a patient experience survey. This particular study attempted to enhance content validity by tailoring the survey to the experiences, needs and preferences of the specific patient group. In

research on quality of care, the experiences of children and (pre)adolescents are usually assessed by asking their parents. However, these young patients may have preferences of their own, and their experiences do not necessarily concur with those of their parents. Our research aimed to give young patients an opportunity to speak up for themselves.

Focus group meetings and online focus groups were organized for two age groups (8–11 and 12–15 years) of young patients from two Dutch rehabilitation centres, with the use of a tailored interview technique. The feasibility and applicability of both types of focus groups were assessed. The results were used to develop a patient experience survey for children and (pre)adolescents on their treatment in rehabilitation centres.

For both types of focus groups, recruitment proved a problem. The focus group meetings with (pre)adolescents proved both feasible and applicable in obtaining information regarding their preferences and experiences towards rehabilitation care. Regarding children's meetings, there were mixed results. The setting suited most of them, but some were somewhat uncomfortable. In general, online focus groups were less successful than expected for both age groups.

Possibilities for facilitating recruitment should be considered, as we may not have reached information saturation from the focus groups, due to the low participation rates. The tailored design proved useful for obtaining relevant input from (pre)adolescents through meetings, but, especially for children, repeated meetings or other locations (e.g. at home) could be considered. This may make participants more at ease. For both age groups, the online focus groups proved far less useful, in terms of participation.

Chapter 3 described the entire development process of a patient experience survey, in this case the CQI Chronic Skin Diseases (CQI-CSD). Although patient experiences with care are of increasing importance in assessing quality of healthcare, to date, no standardized instrument was available in dermatology. The goal of this study was to underpin the validity of the patient experience survey, in development, psychometric testing and optimization. For instance, how are the contents of the survey obtained and which tests are used to investigate its validity and reliability?

This study aimed 1) to evaluate the dimensional structure of the CQI-CSD, 2) to assess its ability to distinguish between hospitals according to patients' experiences with quality of care, 3) to explore patient experiences with dermatological care and priorities for quality improvement according to patients, and 4) to optimize the questionnaire based on psychometric results and stakeholders' input. In a cross-sectional study, 5,647 adult patients who

received dermatological care in the past 12 months in 20 hospitals were randomly selected and invited to fill out the questionnaire.

1,160 of 3,989 eligible respondents (RR 29%, 30 - 87 per hospital) were included for analyses. The CQI-CSD comprised seven reliable scales (Cronbach's α 0.74-0.92). The instrument's discriminative power was limited. Patients were positive about the care provided by nurses and doctors, but the provision of information by healthcare providers, accessibility of care and patient involvement could be improved. Participants were asked to rate the quality aspects included in the survey on their importance. Quality aspects rated as less important were considered for deletion from the survey. The CQI-CSD was optimized according to patient-rated importance and consensus among stakeholders, resulting in a revised questionnaire.

The example of the CQI-CSD demonstrates that by involving all relevant stakeholders (patients, healthcare professionals and health insurers), a valid, reliable and useful survey was developed to measure patient experiences with dermatological care.

In **Chapter 4**, a study was presented on combining patient-reported experience measures (PREMs), patient-reported outcome indicators (PROMs) and healthcare provider-reported clinical indicators. Traditionally, patient surveys for measuring quality of care consisted mainly of patient evaluations of the care process (PREMs), but there is increasing interest to include PROMs. So what do PROMs specifically add to the questionnaires and how do PREMs and PROMs relate to each other? By examining this, we obtained a view on their construct validity. Moreover, this study also assessed the associations between PREMs, PROMs and clinical indicators reported by healthcare providers themselves, thus comparing the perspectives of patients and healthcare providers.

Hospitals were selected for which PREMs, PROMs and clinical indicators were available. 45 hospitals (34% of Dutch hospitals) could be included, involving the experiences of 5,055 patients (RR 63%), of whom 2,720 underwent hip surgery and 2,335 knee surgery.

Associations between PREMs, PROMs and clinical indicator scores were assessed using Pearson correlation coefficients. To this end, PREM and PROM scores were aggregated to the hospital level, adjusted for case mix. We found a few moderate but positive associations between patient experiences (PREMs) and patient-reported outcomes (PROMs) of hip and knee surgery at the hospital level. Some correlations proved significant for the specific knee PROM, but none did for the specific hip PROM. Regarding clinical indicators, only a few significant relationships were observed with PREMs and PROMs. Most notably, the average number of hip or knee arthroplasties carried out

per specialist was positively related to a number of PREMs of both hip and knee surgery.

We found PREMs and PROMs to be complementary in describing quality of care from the patients' perspective. Also, surgeon volume seemed especially positively related to PREMs. Our analyses were to some extent impeded by the number of hospitals included and a limited variation of clinical indicator scores. This study illustrated both the potential benefits and difficulties of combining quality of care indicators from different stakeholder perspectives and/or different data sources.

Patient surveys generally produce numerous results: multiple performance scores – both detailed and aggregated – figures, tables, and so on. Global ratings of healthcare by patients are a popular way of summarizing patients' experiences and are frequently used for comparing healthcare provider performance and for establishing provider rankings. As an alternative, overall scores from actual patient experiences can be applied. **Chapter 5** addressed the statistical and practical characteristics of overall scores as an alternative to a global rating in summarizing patient survey results. An important condition is that such overall scores are representative for actual patient experiences, thus relating to their construct validity.

Data from a 2010 patient experience survey for approximately 12,000 nursing home residents (7.5% of all Dutch nursing home residents at the time) from 464 nursing homes in the Netherlands (25% of the Dutch nursing homes) was used. Data was collected through specifically designed standardized interview surveys (CQI on Nursing Home Care).

The respondents' scores for 15 established quality indicators (or composites) for nursing home care were used to calculate overall scores for each nursing home, using four different strategies. The characteristics of the overall scores were compared to each other and to the respondents' global rating.

Individual quality indicators showed stronger associations with each of the four overall strategies than with a global rating. As a result, overall scores proved more valid than global ratings as a summary of patient experiences. Also, the overall scores showed more pronounced differences between nursing homes. Because of the limited statistical differences between the strategies, and for practical reasons, a straightforward averaging of quality indicator scores may be preferred as an overall score.

In the search for more straightforward ways of summarizing patient experiences and satisfaction, there is growing interest in the Net Promoter Score (NPS): 'How likely is it that you would recommend our company to a

friend or colleague?' **Chapter 6** contained a study on the construct validity of the NPS. Because the NPS is considered to replace the global rating and a recommendation question in CQI surveys as a summarizing measure, their respective relationships were investigated. To establish whether the NPS is a valid measure for summarizing patient experiences, its association with these experiences was assessed.

Data were used from the patient surveys of the CQI Inpatient Hospital Care (N=6,018) and CQI Outpatient Hospital Care (N=10,902) in six Dutch hospitals.

The NPS showed moderate to strong correlations with a global rating, the CQI recommendation question and an overall score of patient experiences, but their distributions proved distinctly different. Also, the NPS reflected the scores of the quality indicators on patient experiences poorly, making it seemingly less valid as a summary of patient experiences than a global rating or, especially, an overall score.

The NPS seems less valid as a summary of patient experiences than a global rating or an overall score calculated from patient experiences. In short, it is unclear what the NPS specifically adds to patient experience surveys.

Another point about the level of detail of survey results is the level at which patient experiences are measured, analysed and presented. The validity and usability of results depend in part on whether the right unit of observation is being described. This issue was addressed in **Chapter 7**, for the case of hospital care. In this setting it may be important to know the quality of care in different departments, in addition to aggregated information at hospital level. This might provide more specific and therefore more useful information.

The aim of this study was to demonstrate the added value of measuring and presenting patient experiences at the department level, in addition to the hospital level, and to explore the possibility that patient experiences differ according to the *type* of hospital department.

Secondary analyses were performed using data from the CQI Inpatient Hospital Care survey. This included the responses of 15,171 inpatients from 78 Dutch hospitals, who had at least one night of hospitalization.

Adding the department level to the analyses of patient experiences is statistically worthwhile for a number of quality indicators of the CQI Inpatient Hospital Care, and will enable the presentation of more detailed results within hospitals. Furthermore, the results indicated that there are some systematic differences in patient experiences between specific types of hospital departments across hospitals. However, the proportion of variance in experiences explained by both department and hospital is limited for most quality indicators.

To conclude, in order to enhance the usability of survey results, quality information on patient experiences of inpatient hospital care should not only be analysed and presented at the hospital level, but also at the more specific department level.

Discussion and conclusions

Validity

A valid patient experience survey measures aspects of care that are relevant to all stakeholders, but especially to patients. As these surveys concern quality of care from their perspective, the patients' view cannot be substituted with that of other stakeholders. They are the most important source of information, both as participants when filling in the surveys and as stakeholders in the process of developing and tailoring these surveys. This is at the core of the content validity of patient experience surveys.

Various qualitative research methods are suitable for identifying aspects of care that are most important to patients. Focus group meetings are a highly popular method to do so. Online focus groups may also be used, although traditional focus groups were more fruitful in our experience. Another possibility is the use of individual interviews. Subsequently, letting patients prioritize quality aspects is an appropriate method.

Patients need to be able to fully report their care experiences in the survey. To this end, survey items need to be phrased in an understandable way, arranged logically, and item response categories need to cover all possible answers from respondents. To accomplish this, a qualitative development phase, cognitive interviewing and extensive testing of (new) survey items are necessary. Otherwise, survey results may prove meaningless in the end.

Validity of patient experience surveys remains an issue, however, since stakeholders have several purposes regarding the outcomes of patient surveys and these purposes may be subject to change. This is exemplified by the addition, deletion or substitution of survey items. These actions do not have to be problematic, but may have consequences for the validity of the survey. To find out which consequences, assessments of response patterns and relationships between survey items may prove their worth, as was illustrated in this thesis.

Usability

Patient survey research produces large amounts of data from which various information products can be made, such as figures, tables, and so on. This information serves different goals. The usability of quality information depends on whether aspects of care relevant to the stakeholders are

measured. But also, whether this information is presented in a way understandable and applicable for these stakeholders.

When it comes to analysing the data and presenting the results, it is often impossible to satisfy the needs of all stakeholders in a single analysis or output format. In this thesis, we have seen some examples of ways to summarize information, which may lead to more manageable results for some, but not for all stakeholders. Some need detailed information, others aggregated information.

Regarding the latter purpose, summary scores are thought to be useful. The use of survey items or overall scores as summary scores of patient experiences is at the cutting edge of both validity of patient surveys and usability of survey results. We examined a number of these scores on their merits. We found a simple averaging of quality indicator scores both a valid and a pragmatic choice, as compared to measures such as the NPS, global ratings and recommendation questions.

Other stakeholders, for instance healthcare providers, have need for more detailed information. They need this information to identify which aspects or processes of their care should be improved, for which patient group, and at what department. This is needed to enable targeted action. Although the influence of institutional and department characteristics on differences in patient experiences proved relatively small, the level of detail is essential to identify where improvements should be made.

Among patients, there are large differences in their information needs; some are highly interested in detailed information, others are content with a general overview of quality of care, and still others are not at all inclined to search or use this information, due to a lack of capabilities (e.g. health literacy), or lack of motivation.

In short, it is important to clearly define the purposes each stakeholder has with regard to the data. Including all relevant stakeholders at an early stage of the research is necessary. In this way, their information needs can be assessed and the data analyses may be adjusted accordingly. Although this sounds reasonable, in practice, use and usability still remain issues. It is not always clear which patients, healthcare providers and health insurers use quality information. And for those who do not use the information, what are their reasons not to do so? The presentation of highly detailed information may lead to confusion, whereas the use of summarizing scores inevitably leads to a simplification of reality.

Ideally, in each research project, it should be considered at an early stage for whom the results are relevant and how to present them in a suitable way. Unfortunately, one size does not seem to fit all stakeholder needs and preferences.

Samenvatting (summary in Dutch)

Patiëntervaringen zijn een belangrijk onderdeel geworden van informatie over kwaliteit van zorg. In Nederland wordt deze informatie gebruikt door drie belanghebbende partijen (of stakeholders) in een systeem van gereguleerde concurrentie: door patiënten voor het kiezen van een zorgaanbieder, door zorgverleners om de zorg te verbeteren en om verantwoording af te leggen over de door hen verleende zorg, en door zorgverzekeraars bij de inkoop van zorg (Enthoven en Van de Ven, 2007). Valide, betrouwbare en bruikbare informatie over kwaliteit van zorg, inclusief het patiëntenperspectief, zou volgens alle stakeholders centraal moeten staan in dit systeem.

Om de ervaringen van Nederlandse patiënten op een meer gestandaardiseerde manier te meten zijn er sinds 2006 patiëntvragenlijsten van de Consumer Quality Index (CQ-index of CQI) ontwikkeld (Delnoij *et al.*, 2006; 2010). Een overzichtsstudie van de onderzoeken uit de eerste vijf jaar van zijn bestaan gaven een tamelijk positief beeld van de CQ-index, maar ook een aantal punten van zorg en mogelijkheden voor verbetering (De Boer *et al.*, 2011a).

Voor wat betreft de surveymethodologie bleken er aandachtspunten te zijn over de validiteit van de patiëntervaringsvragenlijsten en de bruikbaarheid van de resultaten van deze vragenlijsten. De validiteit van de vragenlijsten hangt af van de mate waarin deze de meest relevante aspecten van de kwaliteit van de zorg bevatten (indruks- en inhoudsvaliditeit) en of de vragenlijstitems meten wat ze beogen te meten (constructvaliditeit) (Streiner en Norman, 1999c; Mokkink *et al.*, 2012). De bruikbaarheid van de patiëntervaringsvragenlijsten draait erom of de belanghebbende partijen de resultaten kunnen gebruiken. Zijn ze bijvoorbeeld in staat om de gepresenteerde resultaten goed te interpreteren? En kunnen ze deze resultaten vervolgens gebruiken om een zorgaanbieder te kiezen (patiënten), zorg te verbeteren (zorgverleners) of om kwalitatief goede zorg in te kopen (zorgverzekeraars)?

Dit zijn belangrijke aspecten van het meten en verbeteren van de kwaliteit van zorg vanuit patiëntenperspectief. Daarnaast lijkt het erop dat de informatie over kwaliteitsinformatie over patiëntervaringen niet altijd voldoet aan de eisen van de stakeholders. Dit bevestigt het belang van verder onderzoek om de validiteit van patiëntervaringsvragenlijsten en de bruikbaarheid van vragenlijstresultaten beter te begrijpen en vervolgens te verbeteren. In dit proefschrift werden zes studies gepresenteerd over deze kwesties. Dit proefschrift trachtte twee algemene vragen te beantwoorden:

1. *Hoe kan de validiteit van patiëntervaringsvragenlijsten worden verbeterd?*
2. *Hoe kan de bruikbaarheid van de resultaten van patiëntervaringsvragenlijsten worden verbeterd voor de stakeholders?*

Hoofdstuk 2 richtte zich op de inhoudsvaliditeit van een patiëntervaringsvragenlijst. In deze studie is geprobeerd om de inhoudsvaliditeit te verbeteren door de vragenlijst af te stemmen op de ervaringen, behoeften en voorkeuren van een specifieke patiëntengroep. In onderzoek naar kwaliteit van zorg worden de ervaringen van kinderen en jongeren meestal gerapporteerd door hun ouders. Deze jonge patiënten hebben echter mogelijk eigen voorkeuren en hun ervaringen komen niet per se overeen met die van hun ouders. Ons onderzoek was bedoeld om jonge patiënten de mogelijkheid te geven om hun eigen mening te geven.

Er zijn focusgroepbijeenkomsten en online focusgroepen georganiseerd voor twee leeftijdsgroepen (8-11 en 12-15 jaar) van jonge patiënten in twee Nederlandse revalidatiecentra, met het gebruik van een op maat gesneden onderzoeksdesign. De haalbaarheid en toepasbaarheid van beide typen focusgroepen werden beoordeeld. De resultaten werden gebruikt om een patiëntervaringsvragenlijst voor kinderen en jongeren te ontwikkelen, over hun behandeling in revalidatiecentra.

Voor beide typen focusgroepen bleek de werving een probleem. De focusgroepbijeenkomsten met jongeren bleken haalbaar en goed toe te passen om informatie te verzamelen over hun voorkeuren en ervaringen in de revalidatiezorg. De resultaten van de bijeenkomsten met de kinderen waren echter gemengd. De situatie bleek geschikt voor de meeste kinderen, maar sommigen waren niet zo op hun gemak. Online focusgroepen bleken uiteindelijk minder succesvol dan verwacht voor beide leeftijdsgroepen.

Er zou moeten worden nagedacht over mogelijkheden om de werving voor de focusgroepen te verbeteren, aangezien er door het kleine aantal deelnemers mogelijk geen verzadiging van de verzamelde informatie is bereikt. Het op maat gesneden onderzoeksdesign is nuttig gebleken voor het verkrijgen van relevante input van jongeren door middel van bijeenkomsten, maar vooral voor kinderen zou er nagedacht kunnen worden over herhaalde bijeenkomsten of een andere locatie (bijvoorbeeld thuis). Dit kan ervoor zorgen dat deelnemers meer op hun gemak zijn. Door de beperkte deelname bleken de resultaten uit de online focusgroepen voor beide leeftijdsgroepen weinig bruikbaar.

Hoofdstuk 3 beschreef het volledige ontwikkelproces van een patiëntervaringsvragenlijst, in dit geval de CQI Chronische huidziekten (CQI-CSD). Hoewel patiëntervaringen van toenemend belang zijn bij de beoordeling van de kwaliteit van zorg, was er tot op heden in de dermatologie geen gestandaardiseerd instrument voor beschikbaar. Het doel van deze studie was om de validiteit van de patiëntervaringsvragenlijst te onderbouwen in de ontwikkeling, psychometrische test en optimalisatie. Hoe is bijvoorbeeld de inhoud van de vragenlijst vastgesteld en welke methoden zijn gebruikt om de validiteit en de betrouwbaarheid te onderzoeken?

Deze studie had als doel om: 1) de dimensionele structuur van de CQI-CSD te evalueren, 2) het onderscheidend vermogen te onderzoeken op basis van patiëntervaringen met verschillende ziekenhuizen, 3) patiëntervaringen met dermatologische zorg en prioriteiten voor kwaliteitsverbetering volgens patiënten vast te stellen, en 4) de vragenlijst te verbeteren op basis van de psychometrische resultaten en commentaar van stakeholders. In een cross-sectionele studie werden 5.647 volwassen patiënten uit 20 ziekenhuizen willekeurig geselecteerd en uitgenodigd om de vragenlijst invullen. Deze patiënten hadden in de afgelopen 12 maanden dermatologische zorg gehad.

Het onderzoek bleek van toepassing op 3.989 patiënten. Uiteindelijk konden de antwoorden van 1.160 patiënten (respons 29%, 30-87 patiënten per ziekenhuis) worden meegenomen in de analyses. De CQI-CSD omvatte zeven betrouwbare schalen (Cronbach's α 0,74-0,92). Het onderscheidend vermogen van het instrument was beperkt. De patiënten waren positief over de zorg door verpleegkundigen en artsen, maar er was verbetering mogelijk voor wat betreft het verstrekken van informatie door de zorgverleners, de toegankelijkheid van de zorg en het betrekken van de patiënt. Respondenten werd ook gevraagd om te bepalen hoe belangrijk zij de kwaliteitsaspecten in de vragenlijst vonden. Indien kwaliteitsaspecten als minder belangrijk werden beoordeeld, kwamen deze in aanmerking voor verwijdering uit de vragenlijst. Op basis van de door de respondenten beoordeelde belangscores en consensus onder de stakeholders werd de CQI-CSD geoptimaliseerd. Dit resulteerde in een herziene vragenlijst.

Het ontwikkelproces van de CQI-CSD laat zien dat er door het betrekken van alle belanghebbende partijen (patiënten, zorgverleners en zorgverzekeraars) een valide, betrouwbare en bruikbare vragenlijst is ontwikkeld om de ervaringen van patiënten met dermatologische zorg te meten.

In **hoofdstuk 4** werd een onderzoek gepresenteerd over het combineren van patiëntervaringen met de zorg (*Patient Reported Experience Measures*; PREMs), door patiënten gerapporteerde uitkomstindicatoren (*Patient Reported Outcome Measures*; PROMs) en zorginhoudelijke indicatoren die door zorgverleners gerapporteerd zijn. Van oorsprong bestonden patiëntvragenlijsten over kwaliteit van zorg voornamelijk uit beoordelingen van patiënten van het zorgproces (PREMs). Er is echter een toenemende interesse om PROMs toe te voegen aan deze vragenlijsten. Wat voegen PROMs nu precies toe aan deze vragenlijsten en hoe verhouden PREMs en PROMs zich tot elkaar? Door dit te onderzoeken kregen we een idee van hun constructvaliditeit. Bovendien zijn in deze studie ook de associaties tussen PREMs, PROMs en door zorgverleners gerapporteerde klinische indicatoren onderzocht. Daardoor werden de perspectieven van patiënten en zorgverleners vergeleken.

Voor deze studie werden ziekenhuizen geselecteerd waarvoor PREMs, PROMs en klinische indicatoren beschikbaar waren. De ervaringen van 5.055 patiënten (respons 63%) uit 45 ziekenhuizen (34% van de Nederlandse ziekenhuizen) konden worden meegenomen. 2.720 van deze patiënten ondergingen een heupoperatie en 2.335 een knieoperatie.

De verbanden tussen PREMs, PROMs en klinische indicatorscores werden beoordeeld door middel van Pearson correlatiecoëfficiënten. Daartoe werden de PREM- en PROM-scores geaggregeerd naar het ziekenhuisniveau, gecorrigeerd voor case-mix. We vonden op ziekenhuisniveau een paar bescheiden maar positieve verbanden tussen patiëntervaringen (PREMs) en door patiënten gerapporteerde uitkomsten (PROMs) van de heup- en knieoperaties. Sommige correlaties bleken significant voor de specifieke knie-PROM, maar geen bleek significant voor de specifieke heup-PROM. Ten aanzien van klinische indicatoren werden slechts enkele significante verbanden gevonden met PREMs en PROMs. Het meest opvallend was het positieve verband tussen het gemiddelde aantal heup- of knieervangingen per specialist en een aantal PREM-scores.

PREMs en PROMs bleken complementair in het beschrijven van kwaliteit van zorg vanuit het perspectief van de patiënt. Ook leek het behandelvolume per chirurg positief gerelateerd aan PREMs. Onze analyses werden enigszins belemmerd door het beperkte aantal ziekenhuizen dat kon worden meegenomen en hun beperkte variatie in klinische indicatorscores. Deze studie toonde zowel de mogelijke voordelen als de moeilijkheden van het combineren van indicatoren voor kwaliteit van zorg vanuit verschillende stakeholderperspectieven en/of verschillende gegevensbronnen.

Patiëntvragenlijsten leveren doorgaans veel gegevens op: meerdere prestatiescores – zowel gedetailleerd als geaggregeerd – figuren, tabellen, enzovoort. Een populaire manier om patiëntervaringen samen te vatten is door patiënten rapportcijfers te laten geven aan de zorg. Zulke rapportcijfers worden vaak gebruikt om de prestaties van zorgverleners te vergelijken en om zorgverleners te rangschikken. Als alternatief hiervoor kunnen ook totaalscores van de gemeten patiëntervaringen worden toegepast. **Hoofdstuk 5** richtte zich op de statistische en praktische eigenschappen van totaalscores om de resultaten van patiëntvragenlijsten samen te vatten, als alternatief voor een rapportcijfer. Een belangrijke voorwaarde met betrekking tot constructvaliditeit is dat dergelijke totaalscores representatief zijn voor de werkelijk gemeten ervaringen van patiënten.

Voor de analyses werden gegevens gebruikt uit een vragenlijstonderzoek uit 2010 naar patiëntervaringen met verpleeghuiszorg. De ervaringen van ongeveer 12.000 verpleeghuisbewoners (7,5% van alle Nederlandse verpleeghuisbewoners op dat moment) van 464 verpleeghuizen in Nederland (25% van de Nederlandse verpleeghuizen) werden meegenomen in de analyses. De gegevens werden verzameld door middel van speciaal ontworpen

gestandaardiseerde interviews met de CQ-index Verpleging, Verzorging en Thuiszorg.

De totaalscores voor elk verpleeghuis zijn berekend met de scores van de respondenten op 15 kwaliteitsindicatoren. Hiervoor zijn vier verschillende strategieën gebruikt. De kenmerken van de totaalscores werden vergeleken met elkaar en met de rapportcijfers die respondenten hadden gegeven.

De individuele kwaliteitsindicatoren hielden sterker verband met elk van de vier totaalscores dan met het rapportcijfer. Hierdoor bleken totaalscores een meer valide manier om patiëntervaringen samen te vatten dan het rapportcijfer. Daarnaast lieten de totaalscores meer verschillen zien tussen verpleeghuizen. Als het aankomt op het kiezen van een totaalscore zou een eenvoudig gemiddelde van indicatorscores de voorkeur kunnen krijgen. Dit vanwege de beperkte statistische verschillen tussen de strategieën en om praktische redenen.

In de zoektocht naar eenvoudiger manieren om patiëntervaringen en -tevredenheid samen te vatten, is er een groeiende belangstelling voor de Net Promoter Score (NPS): 'Hoe waarschijnlijk is het dat u ons bedrijf zou aanbevelen bij een vriend of collega?' **Hoofdstuk 6** bevatte een studie over de constructvaliditeit van de NPS. Er wordt overwogen om in CQI-vragenlijsten het rapportcijfer en de aanbevelingsvraag te vervangen door de NPS als een samenvattende maat. Daarom werden in deze studie hun onderlinge relaties onderzocht. Om vast te stellen of de NPS een valide maat is voor het samenvatten van ervaringen van patiënten, werd ook de associatie van de NPS met deze ervaringen beoordeeld.

Er zijn gegevens gebruikt van de patiëntervaringsvragenlijsten van de CQI Ziekenhuisopname (N=6.018) en de CQI Poliklinische ziekenhuiszorg (N=10.902) in zes Nederlandse ziekenhuizen.

De NPS toonde matige tot sterke correlaties met het rapportcijfer, de CQI-aanbevelingsvraag en een totaalscore van patiëntervaringen, maar hun antwoordverdelingen bleken duidelijk verschillend. Bovendien weerspiegelde de NPS de scores van de kwaliteitsindicatoren van patiëntervaringen slecht. Hierdoor lijkt de NPS minder valide als een samenvatting van patiëntervaringen dan een rapportcijfer, en vooral dan een totaalscore berekend op basis van patiëntervaringen. Het is daarmee onduidelijk wat de toegevoegde waarde is van de NPS aan patiëntervaringsvragenlijsten.

Een ander aandachtspunt wat betreft het detailniveau van vragenlijstresultaten, is het niveau waarop de patiëntervaringen worden gemeten, geanalyseerd en gepresenteerd. De validiteit en bruikbaarheid van de resultaten hangen deels af van of de juiste setting wordt onderzocht. Deze kwestie werd behandeld in **hoofdstuk 7**, in dit geval voor ziekenhuiszorg. In deze sector kan het van belang zijn om de kwaliteit van zorg niet alleen op ziekenhuisniveau te kennen, maar ook op het niveau van de verschillende

afdelingen. Dit kan specifiekere en daardoor meer bruikbare informatie opleveren.

Het doel van deze studie was om te laten zien dat het meten en presenteren van patiëntervaringen behalve op ziekenhuisniveau ook op afdelingsniveau van toegevoegde waarde kan zijn. Ook is nagegaan of patiëntervaringen verschillen naargelang het *type* ziekenhuisafdeling.

Er zijn secundaire analyses uitgevoerd met gegevens van de CQI Ziekenhuisopname. Het ging om de antwoorden van 15.171 patiënten van 78 Nederlandse ziekenhuizen, die minstens één nacht opgenomen waren geweest in het ziekenhuis.

Voor een aantal kwaliteitsindicatoren van de CQI Ziekenhuisopname bleek het statistisch gezien de moeite waard om het afdelingsniveau toe te voegen aan de analyses van de patiëntervaringen. Deze toevoeging maakt het mogelijk om meer gedetailleerde resultaten binnen ziekenhuizen te presenteren. Daarnaast bleek uit de resultaten dat er een aantal systematische verschillen in patiëntervaringen was tussen specifieke typen ziekenhuisafdelingen. De verklaarde variantie in ervaringen op afdelings- en ziekenhuisniveau was voor de meeste kwaliteitsindicatoren echter beperkt.

Kortom, om de bruikbaarheid van vragenlijstresultaten te verbeteren zou kwaliteitsinformatie over patiëntervaringen met intramurale ziekenhuiszorg niet alleen moeten worden geanalyseerd en gepresenteerd op ziekenhuisniveau, maar ook op het meer specifieke afdelingsniveau.

Discussie en conclusies

Validiteit

Een valide patiëntervaringsvragenlijst meet aspecten van de zorg die relevant zijn voor alle stakeholders, maar vooral voor patiënten. Aangezien deze vragenlijsten gaan over kwaliteit van zorg vanuit hun perspectief, kan de mening van de patiënt niet worden vervangen met die van andere belanghebbenden. Patiënten zijn de belangrijkste bron van informatie; als deelnemers bij het invullen van de vragenlijst en als belanghebbenden bij het ontwikkelen en bewerken van deze vragenlijsten. Dit raakt de kern van inhoudsvaliditeit van patiëntervaringsvragenlijsten.

Voor het identificeren van zorgaspecten die het meest belangrijk voor patiënten zijn verschillende geschikte kwalitatieve onderzoeksmethoden voorhanden. Focusgroepbijeenkomsten zijn hierbij een zeer populaire methode. Online focusgroepen kunnen ook worden gebruikt, hoewel de traditionele focusgroepen in onze ervaring meer opleverden. Een andere mogelijkheid is om individuele interviews te houden. In de loop van het project kan aan patiënten gevraagd worden om de genoemde kwaliteitsaspecten te prioriteren.

Patiënten moeten in de vragenlijsten hun ervaringen met de zorg zo volledig mogelijk kunnen rapporteren. Daartoe dienen de vragen in de

vragenlijst op een begrijpelijke manier geformuleerd zijn, op een logische volgorde staan en alle antwoordcategorieën bevatten die van toepassing kunnen zijn op de situatie van respondenten. Om dit te bereiken zijn een kwalitatieve ontwikkelingsfase, cognitieve interviews en uitgebreide tests van (nieuwe) vragenlijstitems essentieel. Anders kunnen de resultaten van de vragenlijst uiteindelijk toch zinloos blijken.

De validiteit van patiëntervaringsvragenlijsten blijft echter een punt van aandacht, aangezien de belanghebbende partijen verschillende doelen hebben met betrekking tot de uitkomsten van het vragenlijstonderzoek. Deze doelen kunnen ook aan verandering onderhevig zijn. Dit blijkt uit de voortdurende toevoeging, verwijdering of vervanging van vragenlijstitems. Deze aanpassingen hoeven niet problematisch te zijn, maar kunnen wel gevolgen hebben voor de validiteit van de vragenlijst. Uit dit proefschrift bleek dat het onderzoeken van antwoordpatronen en van relaties tussen vragenlijstitems waardevol kan zijn om te weten welke gevolgen aanpassingen van de vragenlijst kunnen hebben.

Bruikbaarheid

Onderzoek met patiëntvragenlijsten levert grote hoeveelheden gegevens op, waaruit veel informatie kan worden geput, zoals figuren, tabellen, enzovoort. Deze informatie dient verschillende doelen. Voor de bruikbaarheid van kwaliteitsinformatie is het belangrijk dat aspecten van de zorg worden gemeten die voor de stakeholders relevant zijn. Daarnaast moet de informatie zo worden gepresenteerd dat stakeholders het kunnen begrijpen en toepassen.

Wat betreft het analyseren van de gegevens en de presentatie van de resultaten is het vaak onmogelijk om aan de behoeften van alle betrokken partijen te voldoen. In dit proefschrift passeerden een aantal voorbeelden de revue om informatie samen te vatten om zo tot beter hanteerbare resultaten te komen. Voor de ene stakeholder kan dit relevant zijn, voor de andere niet: sommigen hebben gedetailleerde informatie nodig, anderen meer algemene informatie.

Voor dit laatste doel worden samenvattende scores nuttig geacht. Het gebruik van vragenlijstitems of totaalscores als samenvattende scores van patiëntervaringen bevindt zich op het snijvlak van de validiteit van patiëntvragenlijsten en de bruikbaarheid van vragenlijstresultaten. Een aantal van deze scores zijn onderzocht op hun toegevoegde waarde. Een eenvoudig gemiddelde van indicatorscores bleek zowel een valide als een pragmatische keuze, in vergelijking met maten zoals de NPS, rapportcijfers en aanbevelingsvragen.

Andere belanghebbende partijen, zoals zorgverleners, hebben behoefte aan meer gedetailleerde kwaliteitsinformatie. Zij hebben deze informatie nodig om te bepalen welke aspecten of processen van hun zorg moeten worden verbeterd, voor welke patiëntengroep en op welke afdeling. Dit is

nodig om gericht actie te kunnen ondernemen. Hoewel de invloed van ziekenhuis- en afdelingskenmerken op verschillen in patiëntervaringen relatief klein bleek, is het detailniveau van essentieel belang om te bepalen waar verbeteringen nodig zijn.

Patiënten verschillen onderling sterk in hun informatiebehoefte; sommigen zijn zeer geïnteresseerd in gedetailleerde informatie, anderen zijn tevreden met een algemeen overzicht van de kwaliteit van zorg. Weer anderen zijn helemaal niet geneigd om kwaliteitsinformatie te zoeken of te gebruiken. Dit kan te maken hebben met een gebrek aan de benodigde vaardigheden (bijvoorbeeld *'health literacy'*), maar ook met een gebrek aan motivatie.

Kortom, het is belangrijk om duidelijk te hebben welke doelen elke stakeholder heeft met betrekking tot de gegevens. Hiervoor is het belangrijk om in een vroeg stadium van het onderzoek alle belanghebbende partijen te betrekken. Zo kan bepaald worden welke informatie zij nodig hebben en kunnen de data-analyses hierop worden aangepast. Hoewel dit redelijk klinkt, zullen het gebruik en de bruikbaarheid van kwaliteitsinformatie in de praktijk nog steeds aandachtspunten blijven. Het is niet altijd duidelijk welke patiënten, zorgaanbieders en zorgverzekeraars kwaliteitsinformatie gebruiken. En wat zijn de redenen om dat niet te doen, voor degenen die deze informatie niet gebruiken? De presentatie van zeer gedetailleerde informatie kan leiden tot verwarring, terwijl het gebruik van samenvattende scores onvermijdelijk leidt tot een vereenvoudiging van de werkelijkheid.

Idealiter zou bij elk onderzoek in een vroeg stadium moeten worden bepaald voor wie de resultaten relevant zijn en hoe deze op een geschikte wijze gepresenteerd kunnen worden. Helaas lijkt *'one size fits all'* niet op te gaan voor de verschillende behoeften en voorkeuren van de stakeholders.

