Universiteit
Leiden
The Netherlands

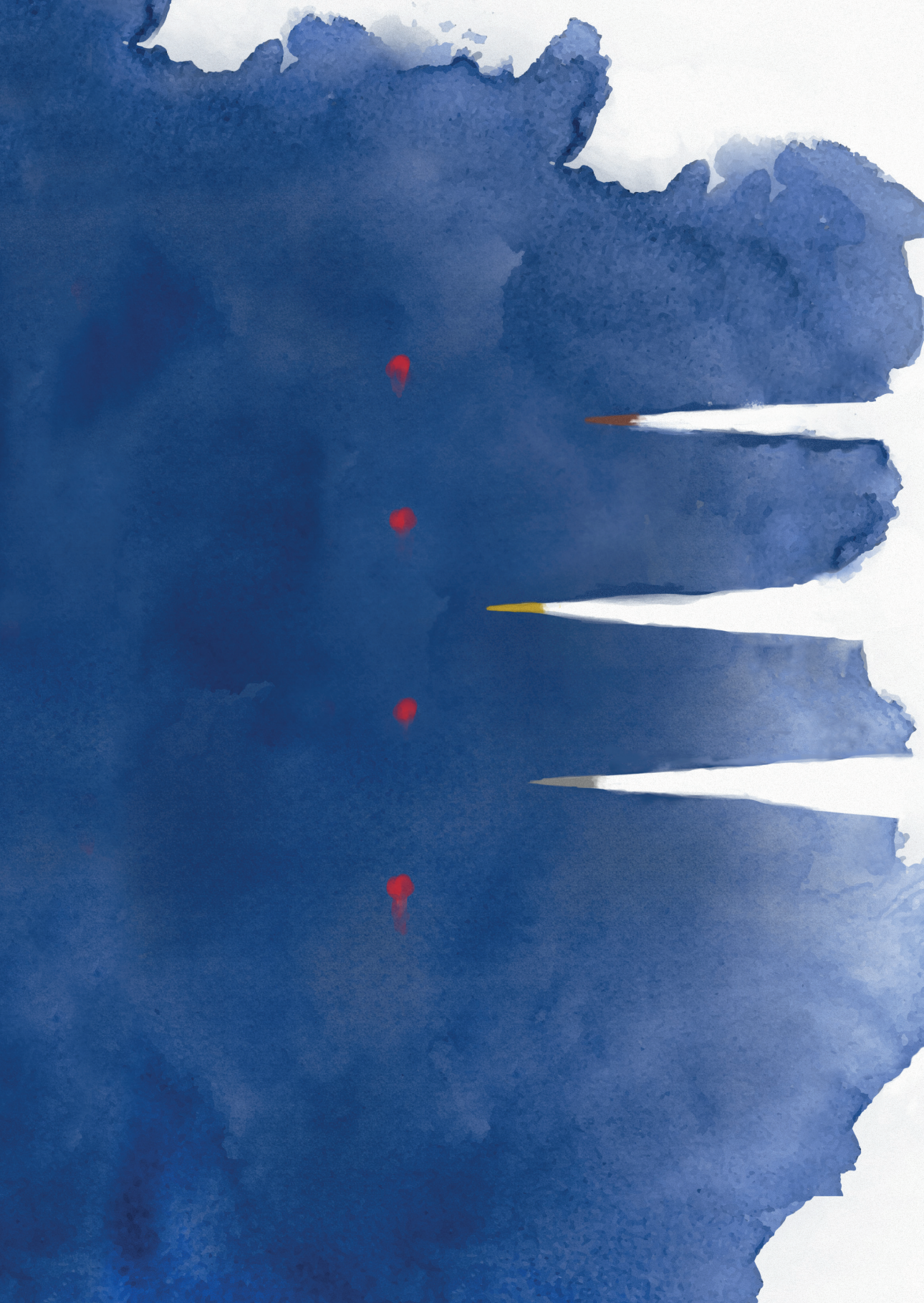
What makes the best performing hospital? the IQ Joint study

Schie, P. van

Citation
Schie, P. van. (2023, November 8). *What makes the best performing hospital?: the IQ Joint study*. Retrieved from https://hdl.handle.net/1887/3656771

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3656771

Note: To cite this publication please use the final published version (if applicable).



Chapter 10

Appendices

Dutch summary (Nederlanse samenvatting)

Het doel van dit proefschrift was om te onderzoeken hoe feedback met betrekking tot totale heup- en knieprotheses (THP en TKP) waarbij gebruik wordt gemaakt van registers kan worden verbeterd. Daarnaast werd de effectiviteit van de verbeterde feedback getest door patiëntenuitkomsten te evalueren.

Protheseregisters werden opgericht om de overleving van implantaten te vergelijken en de veiligheid van verschillende orthopedische implantaten, zoals THP en TKP, te bewaken. De laatste jaren worden registers echter ook gebruikt als kwaliteitssystemen om de variatie tussen ziekenhuizen voor tal van klinische uitkomstmaten inzichtelijk te maken waardoor ziekenhuizen feedback krijgen over hun eigen prestaties. Het laatste wordt meestal gedaan in vergelijking met een referentiestandaard (d.w.z., de benchmark). Het geven van feedback over prestatie-uitkomsten is een veelgebruikte aanpak om de kwaliteit van zorg te verbeteren. In deze context wordt feedback gedefinieerd als het verstrekken van samenvattingen van klinische uitkomsten aan zorgverleners of organisaties met als doel verbeterinitiatieven te initiëren. Internationaal wordt feedback vanuit protheseregisters op verschillende manieren aangeboden. In Nederland worden klinische uitkomsten op ziekenhuisniveau getoond in een actueel dashboard van de Landelijke Registratie Orthopedische Interventies (LROI). Daarnaast wordt de mate van variatie in uitkomsten tussen ziekenhuizen getoond in een geanonimiseerde vorm in jaarlijkse rapporten. Uit de literatuur is gebleken dat de effectiviteit van feedback varieert (d.w.z., van een verslechtering van 9% tot een verbetering van 70%). Een optimale inrichting van de feedback zal dus redelijkerwijs leiden tot een verbetering van de zorg voor de patiënt. Dit proefschrift gaf een overzicht van de nationale en internationale variatie tussen ziekenhuizen in klinische uitkomsten om te onderzoeken of verbetering haalbaar is. Om de feedback inhoudelijk te verbeteren, zijn methodologische studies uitgevoerd om te onderzoeken of zogenaamde uitschieter-ziekenhuizen (zowel positief als negatief) eerder kunnen worden gedetecteerd, of redenen voor hogere revisiepercentages kunnen worden geïdentificeerd, en een samengestelde uitkomstmaat werd ontwikkeld en getest. De verkregen kennis uit deze studies in combinatie met hedendaagse theorieën voor het geven van effectieve feedback werden opgenomen in een veelzijdige kwaliteitsverbeterinterventie waarbij de effectiviteit werd onderzocht in een cluster gerandomiseerde en gecontroleerde trial.

Deze Nederlandse samenvatting beschrijft de belangrijkste bevindingen van dit proefschrift inclusief de praktische implicaties.

Samenvatting en praktische implicaties

Registers geven feedback bedoeld om kwaliteitverbeterinitiatieven in slechter presterende ziekenhuizen aan te jagen. Echter kunnen de best presterende ziekenhuizen ook geïnteresseerd zijn in het vergelijken van hun uitkomsten met ziekenhuizen uit andere landen om verder te verbeteren binnen specifieke domeinen. Eerlijke internationale vergelijking van ziekenhuizen is echter alleen haalbaar wanneer consistente uitkomstdefinities worden gebruikt, omdat deze in grote mate de frequentie bepalen. Consistentie in de definitie van uitkomstmaten maakt het ook mogelijk om internationale gegevens samen te voegen, waardoor betere mogelijkheden ontstaan om zeldzame veiligheidsproblemen eerder te detecteren zoals bijvoorbeeld de problematiek rondom metalen-op-metalen heupprothesen of modulaire femurhalscorrosie, wat duizenden patiënten zal beschermen tegen slecht presterende implantaten, waardoor onnodig lijden in de toekomst wordt verminderd. **Hoofdstuk 2** toonde aan dat revisie, heropname en complicaties de meest gerapporteerde klinische uitkomsten zijn in cohortstudies en rapporten van protheseregisters. Er zijn aanzienlijke verschillen in uitkomsten tussen ziekenhuizen, wat wijst op een groot verbeterpotentieel. Een deel van deze variatie kan echter worden verklaard door de aanzienlijke heterogeniteit in de volgende domeinen: 1) uitkomstdefinities, inclusief wat een revisie, heropname of complicatie inhoudt, 2) duur van follow-up en startpunt van follow-up, 3) kenmerken die zijn opgenomen in de correctie voor de patiëntenmix, en 4) type patiënten en ziekenhuizen die zijn meegenomen in de rapportages. Dit proefschrift toonde aan dat revisie van het implantaat binnen vijf jaar, heropname binnen 30 dagen en complicaties tot 2 jaar na de operatie de meest gebruikte uitkomstmaten waren. Geen van deze definities kwam echter perfect overeen met de andere domeinen voor THA, TKA en THA&TKA gecombineerd. De minste consensus werd gevonden voor welke patiëntkenmerken moet worden gecorrigeerd. Laatstgenoemde evenals de andere onderzochte domeinen zijn essentieel voor een eerlijke vergelijking tussen ziekenhuizen. In de toekomst kunnen samenwerkingen tussen protheseregisters zoals geïnitieerd door de International Society of Arthroplasty Registries, een leidende rol spelen in het streven naar meer uniformiteit in de gebruikte definities en methoden.

Het rapporteren van variatie tussen ziekenhuizen en het identificeren van positieve en negatieve uitschieter-ziekenhuizen wordt gedaan met behulp van funnelplots en is een eenvoudige en effectieve manier om inzicht te krijgen in de prestaties van ziekenhuizen, op voorwaarde dat deze analyses betrouwbaar zijn. De betrouwbaarheid kan worden uitgedrukt als het percentage van de totale variatie dat wordt verklaard door "echte" ziekenhuisverschillen versus toevalsvariatie. **Hoofdstuk 3** toonde een grote variatie in het 1-jaars revisiepercentage tussen Nederlandse ziekenhuizen, met een acceptabele betrouwbaarheid (61%) voor THP en een lage betrouwbaarheid (46%)

voor TKP in een periode van 3 jaar. Dit geeft aan dat er veel verbeteringspotentieel is voor een aanzienlijk aantal Nederlandse ziekenhuizen. Het vroegtijdig opsporen van slechte prestaties met behulp van een tijdspanne van 1 jaar heeft het voordeel dat kwaliteitverbeterinitiatieven eerder kunnen worden geïntroduceerd, maar dit leidde tot een lage betrouwbaarheid en wordt daarom niet aanbevolen. Het beoordelen hoe verbetering voor weinig gespecificeerde uitkomsten zoals 1-jaarsrevisie kan echter uitdagend zijn. Dit kan worden vergemakkelijkt door specifieke redenen voor revisie te onderzoeken die de onderliggende oorzaak kunnen zijn voor slechtere prestaties voor revisie. **Hoofdstuk** 3 toonde dat onderliggende redenen werden gevonden voor 12 van de 13 negatieve uitschieter-ziekenhuizen voor THP en 3 van de 7 voor TKP. De voornaamste onderliggende redenen voor slechtere prestaties waren infectie (zowel voor THP als TKP) en prothesedislocaties (alleen THP). Implant loslating (zowel voor THP als TKP) en technisch falen (alleen TKP) waren minder waarschijnlijk. De betrouwbaarheid voor de specifieke redenen voor revisie waren allemaal laag binnen een tijdspanne van 3 jaar voor THP en TKP, behalve voor infectie voor THP, waarvoor de betrouwbaarheid acceptabel was (d.w.z., 61%). Aangezien de betrouwbaarheid binnen een tijdspanne van 1 jaar allemaal laag waren, wordt aanbevolen om een periode van 3 jaar te gebruiken om de onderliggende redenen voor slechtere ziekenhuisprestaties op het gebied van revisie te identificeren.

In **hoofdstuk 3** werd aangetoond dat het gebruik van funnelplots voor het vroegtijdig opsporen van slechtere prestaties niet betrouwbaar was binnen een tijdspanne van 1 jaar, daarom werd een tijdspannen van 3 jaar aanbevolen. **Hoofdstuk 4** toonde echter aan dat maandelijks monitoren van 1-jaar revisiepercentages met behulp van CUSUM-grafieken met 5 controlelimieten eerder een betrouwbare achteruitgang of verbetering kan detecteren dan de conventionele funnelplots. Het eerste signaal voor negatieve afwijkingen werd gegenereerd na een mediane tijd van 18 maanden voor THP en 21 maanden voor TKP binnen een tijdspanne van 3 jaar. CUSUM-grafieken maken het daardoor mogelijk om eerder een verslechtering te detecteren, waardoor het mogelijk wordt om kwaliteitsverbeterinitiatieven eerder te introduceren dan te wachten op de resultaten in de funnelplot na 3 jaar. De nauwkeurigheid van deze CUSUM-grafieken in vergelijking met de traditionele funnelplot met een tijdspanne van 3 jaar was 97%. Deze resultaten zijn relevant voor protheseregisters en wetenschappelijke verenigingen in de besluitvorming of zij CUSUM-grafieken willen implementeren in hun organisatie. De resultaten over de nauwkeurigheid zullen bijdragen aan het vertrouwen van professionals in CUSUM-grafieken. Naar aanleiding van deze bevindingen heeft de LROI CUSUM-grafieken toegevoegd aan hun routinematige dashboardrapportage voor klinische uitkomsten.

Een andere manier om afwijkende prestaties eerder te kunnen detecteren is het vergrote van het aantal gebeurtenissen zoals bij samengestelde uitkomsten wordt gedaan. In **hoofdstuk 5** werd een geordende samengestelde uitkomst van klinische relevante uitkomsten (d.w.z., revisie, heropname, complicaties en verlengde opnameduur) voor THP en TKP ontwikkeld en getest op de mogelijkheid om nauwkeuriger en eerder tussen ziekenhuizen te differentiëren. Deze nieuwe uitkomst toonde een hogere mate van betrouwbaarheid dan de individuele uitkomsten vanwege de grotere variatie tussen ziekenhuizen wanneer de vier uitkomsten worden samengevoegd. De samengestelde uitkomst kon betrouwbaar verschillen tussen ziekenhuizen in hun prestaties differentiëren met een tijdspanne van 1 jaar, in plaats van de gebruikelijke 3 jaar, waardoor de introductie van kwaliteitverbeterinitiatieven eerder mogelijk wordt. Een bijkomend voordeel is dat de samengestelde uitkomst meerdere aspecten van de geleverde kwaliteit van zorg kan meten, omdat ziekenhuizen het op één uitkomst goed kunnen doen terwijl ze tegelijkertijd slechter kunnen presteren op een andere. De samengestelde uitkomst overwint dit probleem en laat zien of een patiënt een revisie heeft gehad, maar ook of ze opnieuw zijn opgenomen, complicaties hebben gehad of een verlengd opnameduur hebben gehad. Deze nieuwe samengestelde uitkomst kan helpen bij het selecteren van specifieke patiëntengroepen voor medische dossieronderzoek. In plaats van alle dossiers te bekijken voor patiënten die opnieuw zijn opgenomen, maakt de samengestelde uitkomst het mogelijk om selectief bijvoorbeeld alleen diegenen te bekijken die naast een heropname een normale opnameduur en geen complicatie hadden. Zo kan eenvoudiger worden bepaald of bijvoorbeeld informatie bij ontslag moet worden verbeterd om een heropname te voorkomen of dat een heropname mogelijk niet nodig zou zijn geweest indien een patiënt poliklinisch nauwlettender zou zijn vervolgd. Daarnaast is in één uitkomst zichtbaar of een focus om bijvoorbeeld de verblijfsduur te verkorten, niet ten koste gaat van een andere uitkomst zoals bijvoorbeeld heropname. Tot slot kunnen patiënten eenvoudig controleren hoe vaak een procedure verloopt zoals gepland (d.w.z., zonder dat een van de klinische uitkomsten zich voordoet).

Klinische uitkomsten zoals revisie, heropname, complicaties en opnameduur meten onbedoelde gebeurtenissen en komen bij THP en TKP over het algemeen voor met een lage frequentie. Echter 10% van de TKP- en 20% van de THP-patiënten zijn op de lange termijn ontevreden over het resultaat, voornamelijk vanwege aanhoudende pijn en functionele beperkingen. In tegenstelling tot klinische uitkomsten meten Patient-Reported Outcome Measures (PROMs) de bedoelde uitkomsten zoals pijnvermindering, verbetering van functionaliteit en gezondheid gerelateerde kwaliteit van leven en kunnen daarom leiden tot aanvullende verbeterpunten. Net als bij klinische uitkomsten is het van belang dat PROMs data een hoge mate van compleetheid hebben ter voorkomen van selectiebias. In Nederland en andere nationale en regionale

protheseregisters zijn responspercentages echter laag. Dit wil zeggen dat in Nederland 61% van de patiënten de preoperatieve PROM-vragenlijst heeft ingevuld en slechts 40% van de patiënten zowel de pre- als postoperatieve vragenlijst. In tegenstelling tot de compleetheid in registers voor klinische uitkomsten die boven de 98% bedraagt. Eerdere studies hebben al aangetoond dat patiënten die de vragenlijsten invullen over het algemeen gezonder zijn, vaker een witte huidskleur, minder vaak analfabeet en minder vaak cognitieve stoornissen hebben (inclusief dementie). Om beter te begrijpen of ontbrekende PROM-data leidt tot een onder- of overschatting van de postoperatieve PROM-verbeterscores werd in **hoofdstuk 6** het verschil in klinische uitkomsten (d.w.z. revisie, heropname, complicaties en opnameduur) voor respondenten en niet-respondenten beschreven. Ook werd de associatie met PROM-verbeterscores bepaald. PROM-respondenten hadden gunstigere klinische uitkomsten dan niet-respondenten voor THP en voor revisie en heropname voor TKP. THP-patiënten hadden minder kans op een klinisch relevante PROM-verbetering wanneer zij een revisie, complicaties of een verlengde opnameduur hadden met resultaten voor heropname in dezelfde richting maar niet significant. Vergelijkbare resultaten werden gevonden voor TKP. Doordat patiënten die PROM-vragenlijsten invulden over het algemeen gunstigere klinische uitkomsten hadden dan niet-respondenten werd een deel van de patiënten die een kleinere kans hadden op een klinisch relevante verbetering voor de PROMs gemist. PROM-verbeterscores worden daarom waarschijnlijk overschat. Gezien de beschreven associaties in **hoofdstuk 6**, is het waarschijnlijk dat initiatieven om de kwaliteit van zorg te verbeteren door heropname, complicaties en opnameduur voor zowel THP en TKP te verminderen, zullen leiden tot meer patiënten die een klinisch relevante PROM-verbeteringen. Het is onwaarschijnlijk dat ziekenhuisverschillen worden beïnvloed, aangezien de PROM-responspercentages van ziekenhuizen niet geassocieerd waren met klinische uitkomsten.

Feedback is alleen effectief wanneer deze wordt bekeken en correct wordt geïnterpreteerd door de doelgroep. **Hoofdstuk 7** laat zien dat slechts de helft (55%) van de orthopedisch chirurgen in Nederland die THP en TKP plaatsen weet of zij werkzaam zijn in een ziekenhuis die normaal presteert of een positieve of negatieve uitschieter. Het bewustzijn was hoger onder chirurgen die vaker inlogden op het LROI-dashboard, vaker funnelplots correct interpreteerden en vaker de 1-jaars revisiepercentages van hun vakgroep konden herinneren. Achtendertig procent van de THP- en 26% van de TKP-chirurgen voldeed aan deze drie voorwaarden die nodig zijn om actie te ondernemen op basis van de feedbackinformatie. Vijfenveertig procent van de chirurgen in een ziekenhuis dat als negatieve uitbijter was geïdentificeerd, meldde dat ze de verslechterende prestaties van hun ziekenhuis niet zagen aankomen. Dit betekent dat ze doorgingen met het verlenen van zorg zonder wijzigingen aan te brengen. Dit onderbouwt dat het van belang is om te focussen

op het effectiever maken van feedback, mede omdat 85% van de chirurgen aangaf dat er in hun ziekenhuis kwaliteitsverberterinitiatieven werden geïmplementeerd zodra ze werden geïdentificeerd als een slechter presterend ziekenhuis. De chirurgen gaven aan dat inloggen op het LROI-dashboard kan worden aangemoedigd door de inhoud aantrekkelijker te maken, bijvoorbeeld door niet alleen de nadruk te leggen op revisiepercentages, maar nieuwe uitkomsten toe te voegen zoals protheseoverleving, complicaties, heropnames en opnameduur. Deze nieuwe uitkomsten worden namelijk als relevant beschouwd door een groot deel van de chirurgen. Ook zou er onderwijsmateriaal beschikbaar moeten worden gesteld om interpretatievaardigheden van statistische grafieken zoals bijvoorbeeld de funnelplots en CUSUM beter te kunnen interpreteren. Een uitleggende tekst onder statistische grafieken zou ook kunnen helpen om chirurgen de helpen bij het interpreteren van de grafieken. Tot slot zou feedback in een kant en klaar model voor individuele ziekenhuizen (bijvoorbeeld in de vorm van een infographic) moeten worden toegestuurd via de email zodat chirurgen niet zelf selecties moeten maken in een online dashboard en feedback laagdrempelig toegankelijk is.

In **hoofdstuk 8** werden de bevindingen uit de voorgaande hoofdstukken gecombineerd met hedendaagse theorieën over effectieve feedback om een kwaliteitverbeterinterventie te ontwikkelen en de effectiviteit te testen in een cluster gerandomiseerde en gecontroleerde trial. De interventie werd gedurende acht maanden toegepast en omvatte maandelijkse feedback, interactieve educatie over de interpretatie van de feedback en een toolbox met daarin op bewijs gebaseerde kwaliteitverbeterinitiatieven voor alle uitkomsten die in de feedback werden gerapporteerd. Ziekenhuizen die de interventie ontvingen, verbeterden met 4,3% meer op de Textbook Outcome (dat wil zeggen het ontbreken van revisie, heropname, complicaties en een lange opnameduur) in vergelijking met de controlegroep. Het effect was groter bij interventieziekenhuizen die kwaliteitsverbeterinitiatieven introduceerde dan ziekenhuizen die dit niet deden, wat suggereert dat deze kwaliteitsverbeterinitiatieven die werden ondernomen waarschijnlijk de reden waren voor de betere resultaten. Deze bevindingen ondersteunen dat frequente feedback aan chirurgische teams moet worden aangevuld met interactieve educatie en toolbox met daarin kwaliteitverbeterinitiatieven die zijn afgestemd op specifieke uitkomsten om de kwaliteit van zorg met betrekking tot THP en TKP effectief te verbeteren.