Teamwork by surgeon teams in hospitals

An empirical research on the effect of surgeon team size on patient-reported quality of care

Youri Hoppema – 484176
Date: June 23, 2021
Supervisor: Raf van Gestel
Location: Rotterdam
Word count: 10358

Abstract

The fact that health care can both harm and heal patients is reason enough to state that patient safety is at the heart of health care. In modern medicine, health care is often delivered by teams, rather than by individuals. This thesis considers teamwork of surgeons in surgeon teams. Among others, peer learning and efficiency gains can explain part of the success of working in these surgeon teams. However, teamwork may also raise issues with, for example, coordination of care and communication in teams with a larger team size, which may threaten patient safety and adversely affect quality of care. To get insights in the predictors of surgeon team size and in the actual effect of surgeon team size on patient-reported quality of care, a propensity score matching analysis is conducted (N=72). In the analysis I included the difference in EQ-5D index score 12 months after surgery as outcome variable and a dummy variable for hospitals with a large surgeon team size (7 or more surgeons) as dependent variable. I included the following hospital characteristics as covariates: staff-to-bed ratio, the number of procedures per surgeon per year, the OOR region and the type of hospital. My results show a positive and nonsignificant ATT for hospitals with a large surgeon team size on the difference in EQ-5D index score 12 months after surgery. Furthermore, the results show two significant predictors for the indicator of hospitals with a large surgeon team size in the regression. These predictors are surgery volume for TKPs and the indicator for OOR region LUMC for THPs. However, biased results and a poor propensity score distribution make drawing reliable and valid conclusions from these results impossible. I therefore conclude that future research is needed to gain reliable and valid insights in the effects of surgeon team size on patient-reported quality of care.

Table of contents

1. Introduction

The fact that health care can both harm and heal patients is reason enough to state that patient safety is at the heart of health care (Vincent, 2011). In modern medicine, health care is often delivered by teams, rather than by individuals. Physicians cooperate with nurses and anesthesiologists in operating rooms (Makary et al., 2006), primary care physicians cooperate in group practices (Welch et al., 2013) and surgeons work together in surgeon teams. This thesis focuses on this last type of teamwork: the hospital-wide surgeon teams which consist of surgeons only. Henceforth, I will refer to this as "surgeon team", not to be confused with the team in the operating room. Among others, peer learning and efficiency gains can explain part of the success of working in these surgeon teams. However, teamwork may also raise issues with, for example, coordination of care and communication in teams with a larger team size (Makary et al., 2006; Salas et al., 2008).

Evidence from different settings, such as the military, indicates that effective teamwork improves productivity and performance (Paris et al., 2000). These different settings have in common that teamwork in multi-person systems is a central theme. Despite the differences between health care and other settings, evidence from these other environments can still indicate that teamwork in health care is beneficial since health care is delivered through teamwork in multi-person systems too. Surgeons who are working together in a team (effectively) can lead to all kinds of performance gains. For instance, Mickan and Roger (2000) found that teamwork facilitates peer learning with better health care quality as a result. At the same time, teamwork in surgeon teams has been shown to enhance job satisfaction. Besides that, Chan (2016) and Agha et al. (2019) mention that teamwork and other human resource management technologies are associated with higher productivity. Furthermore, another study found that "in terms of delivery of care, teams have been reported to reduce hospitalisation time and costs, improve service provision, enhance patient satisfaction, staff motivation and team innovation" (Borrill et al., 2000).

There are also downsides to teamwork according to Makary et al. (2006). For instance, Makary et al. (2006) argue that communication errors in health care teams are the most common cause of adverse outcomes such as unexpected deaths in the US. Another example is given by Salas et al. (2008), in which the authors define communication, coordination and cooperation as essential teamwork components that are known to be successful in health care. These fundamental requirements of teamwork are interdependently and Salas et al. (2008) describe that "effective coordination requires effective communication, which cannot be effective without cooperating team members". In other words, if one of these elements is not performed effectively, teamwork is not effective and can lead to a performance loss (Salas et al., 2008).

The fact that performance loss among health care teams still exists indicates that there is still opportunity to improve safety and effectiveness in health care. This thesis therefore aims to better understand and contribute to the process of improving health care by looking into the effect of surgeon team size on quality of care.

This thesis considers teamwork by orthopedic surgeons that place a total hip prosthesis (THP) and a total knee prosthesis (TKP). These procedures are two of the most performed procedures in the Netherlands with approximately 26,000 TKPs and 33,000 THPs in 2019, according to the dataset that is used in the analysis. Besides, the actual future prevalence of arthroplasty may also be higher due to the aging of the population and the expected rise in the number of severely overweight people (Tilbury-Werkhoven, 2018).

1.1 Research question

To look into possible associations in which the quality of care is affected by surgeon team size, this thesis analyzes the teamwork of orthopedic surgeons in Dutch hospitals. More specifically, I explore whether there is an association between surgeon team size and (patient-reported) quality of care. This potential association is based on the hypothesis that quality of care is negatively affected by a bigger surgeon team size, because a bigger team size could trigger multiple problems (e.g., communication, cooperation and coordination problems). Other possible mechanisms will be studied in the thesis. Insights into the performance and functioning of surgeon teams in Dutch hospitals are relevant since the current literature lacks this kind of research. The results of this thesis could contribute to evaluating and improving the teamwork of surgeon teams in Dutch hospitals and by giving insights into the effect of surgeon team size on the patient-reported quality of care. Therefore, these results can contribute to patient safety and quality of care in these organizations.

The following research question will be studied:

- *How does hospital-wide orthopedic surgeon team size for THP or TKP procedures affect (patient-reported) quality of care?*

To answer this research question, the following sub research questions are formulated:

- Is there an association between surgeon team size and patient-reported quality of care?

- What are predictors of a hospital's surgeon team size?

2. Theoretical framework

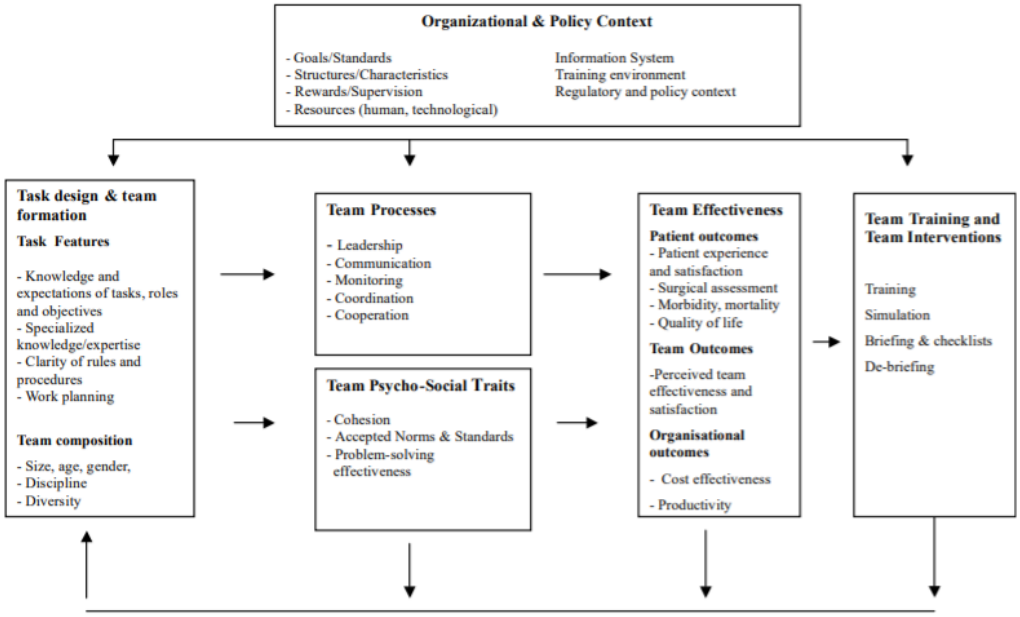
This section discusses the relevant literature on teamwork of surgeon teams and teamwork in other settings. These articles support the understanding of the potential effects of surgeon team size on the quality of care. There is little evidence available on teamwork in hospital-wide surgeon teams, but there is evidence on the mechanisms of teamwork in both health care delivery and other environments. The lessons learned from teamwork in other settings, such as the operating room (OR), can help guide the progression of safer team-driven care in surgeon teams. These lessons learned from both other settings and the mechanisms of health care teams are the main source for this framework.

First, a conceptual model of the effectiveness of surgical teams is outlined. The purpose of this model is to provide insights into factors that influence the performance of surgical teams in the operating room, not to be confused with the surgeon teams of this thesis. These insights provide starting points for this thesis to explore potential causal effects or associations. Following this conceptual model, theoretical perspectives and previous empirical research are provided, which could help in explaining and understanding potential differences in patient-reported quality of care between surgeon teams with different team sizes.

2.1 Theoretical perspective, conceptual models and previous empirical research

The conceptual model of Healey and colleagues (2006) describes multiple interdependent factors that influence surgical team performance and therefore could explain potential differences in team performance between surgical teams in hospitals. The model describes task design and team formation, also known as “input factors”. Vincent (2011) defines team formation as follows: “this refers to the team set-up and concerns who is in the team, what it is meant to be doing, how much autonomy it has and the rules and standards by which the team operates”. In the model, team formation consists of the following team characteristics: age, gender, discipline, diversity and team size. The focus of this framework will be on the effects of team size on other factors in the model and on potential predictors of team size that could be relevant for this thesis.

Figure 1: The surgical team effectiveness model by Healey et al. (2006)



I first discuss the team processes and team psycho-social traits. Vincent (2011) divides team processes into “leadership, communication, monitoring, coordination and cooperation”. Psycho-social traits are traits such as cohesion, the accepted norms and standards and the problem-solving effectiveness. A relationship between team size and both team processes and team psycho-social traits can be observed in the model. This relationship can be translated into multiple hypotheses, indicating that team size affects team processes and psycho-social traits. For example, a change in team size results in differences in team coordination and team communication, because larger teams have problems with coordination and communication (Liang et al., 2008).

Following these communication problems are problems with peer learning. Peer learning is a major benefit of teamwork, because peer feedback within teams can increase the team’s overall efficacy and performance (Salas et al., 2008). Underlying problems, such as non-positive communication among team members, can undermine the principle of peer learning because Salas et al. (2008) argue that “critiques should always be constructive, performance-related, and presented in a nonthreatening way”. As a result, the peers will for example not learn from mistakes that others made or implement new best practices that others did. These are missed opportunities to increase the team’s overall efficacy and performance. This may results in a lower (patient-reported) quality of care for hospitals with a larger surgeon team size in comparison with hospitals with a smaller surgeon team size, because this phenomenon is more likely to occur in hospitals with a larger surgeon team size due to the bigger number of surgeons that are working together. Moreover, a change in team size could change the cohesion in a team too, because smaller teams develop group cohesion more quickly (Thompson et al., 2015). Based on these findings, effects between team size, team processes and psycho-social traits can be assumed.

Furthermore, team effectiveness is affected by team processes and psycho-social traits in the model. Since team size affects both team processes and psycho-social traits, it indirectly affects team effectiveness too. This suggests that there is an indirect relationship between team size and team effectiveness. For example, larger surgeon team sizes in hospitals could result in coordination problems. According to the conceptual model, these coordination problems will affect the team’s effectiveness.

The conceptual model also displays the factors “team training and team interventions” and “organizational context and policy context”. According to Vincent (2011), the team interventions and team training represent “a feedback loop in which clinical outcomes and experiences within the team can all influence subsequent team performance”. This suggests that all factors either directly or indirectly affect team formation and therefore team size. Another factor that is included in the model is organizational context and policy context. This is an independent factor, since it is not affected by other factors in the model. These two types of context come with the working environment of the team and are therefore context specific. These two factors represent influences from factors within these two contexts. An example of a factor in the policy context is the volume-outcome effect. The volume-outcome effect is based on a volume-based policy that introduced minimum volume standards. Mesman et al. (2017) argue that “this policy is aimed at directing certain surgical procedures away from low-volume providers to reduce patients’ risks of unwanted outcomes” and that “since the introduction of these minimum volume standards, researchers have studied the effects of increased surgery volume on health care outcomes”. Surgery volume is defined in this thesis as the number of procedures a surgeon performs per year. Moreover, Mesman et al. (2017) concluded that all included studies in their literature search “show a beneficial effect of increased volume on patient outcomes, while the overall effect on patient welfare and quality of the health care system as a whole remains unclear”. This volume-outcome effect might have an impact on the functioning of orthopedic teams too, because a concentration of procedures in fewer hospitals might for example cause a bigger workload for the orthopedic surgeons and therefore can cause a need for larger surgeon teams in these hospitals, which may affect quality of care.

Various studies have assessed whether surgery volume affects patient outcome positively as well. Katz et al. (2001) found that “patients treated at hospitals and by surgeons with higher annual caseloads (a higher surgery volume) of primary and revision total hip replacement had lower rates of mortality”. Furthermore, a recent study found that “the quality of care in low-volume hospitals is lower than in high-volume hospitals (Kruse et al., 2019). However, Kruse et al. (2019) also state that “the size of this effect is small, and at higher volumes the marginal benefits (in terms of lower postoperative infections) decreases”. Despite the small effect size, this effect is worth taking into account in explaining potential differences in patient-reported outcomes and in understanding the effect of surgeon team size on these outcomes, because the volume-outcome policy is a possible explanatory factor of the change in surgeon team size of hospitals that were affected by minimum volume standards.

Following the results from the studies that are mentioned above, there is enough evidence to state the following two hypotheses:

H1: There is a significant relationship between a hospital’s surgeon team size and the hospital’s patient-reported quality of care.

H2: There is a significant positive relationship between surgery volume and a hospital’s surgeon team size.

This first hypothesis counters the hypothesis that I used to draft the research question of this thesis, in which I stated that there is a negative association between surgeon team size and patient-reported quality of care. To get insights in the predictors of surgeon team size and in the actual effect of surgeon team size on patient-reported quality of care, a propensity score matching analysis is conducted. The following sections describe the research methods and results of this analysis.

3. Research methods

3.1 The propensity score matching analysis

A quantitative analysis evaluates the relationship between the EQ-5D index outcome indicator for THPs and TKPs, and surgeon team size. With the evaluations in this thesis, I aim to better understand and contribute to the process of improving health care by looking into the effect of surgeon team size on quality of care. To do this, a propensity score matching (PSM) analysis is executed with the statistical analysis software STATA to compare surgeon teams of hospitals on multiple outcome indicators. I chose this study design because a big advantage of PSM is that, given that the matching quality is good enough, a propensity score matching analysis allows the estimation of a balancing propensity score which will result in a similar distribution of the observed baseline characteristics between the intervention group and the control group (Austin, 2010). Another advantage of PSM is that it only requires data from one year to compare the observations.

I estimate the effect of being a hospital with a large team size in comparison to a hospital with a small surgeon team size, based on a propensity score. In the PSM model, the mean scores on the EQ-5D index between treated (large surgeon team size hospitals) and control group (small surgeon team size hospitals) observations are compared. With extra information on the characteristics of each hospital, I aimed at finding a group of large surgeon team size hospitals that is comparable to a group of small surgeon team size hospitals. To do this, I created a binary variable that determines if the hospital is a hospital with a large surgeon team size or a hospital with a small surgeon team size. By comparing the hospitals on the EQ-5D outcome indicator, potential relationships between surgeon team size and quality of care are explored. The propensity score matching analysis can be divided in four steps: 1) estimating the propensity score, 2) propensity score matching, 3) evaluating quality of matching, 4) evaluating the outcomes (Pan & Bai, 2015). These steps are discussed in the following four subsections.

First, the estimation of a probit model for the propensity of the observations to be in the treatment group (Rosenbaum & Rubin, 1983). In this thesis, the propensity score is the predicted probability of being a large surgeon team size hospital, given certain covariates of a hospital. These covariates affect the likelihood of being a hospital with a large surgeon team size. A complete list of these covariates is included in Appendix D. After the estimation of the propensity score, the propensity score matching method is determined.

A book on the use of different propensity score analysis methods describe that “there are a number of different propensity score matching methods available that can be used to match units on their propensity scores” (Pan & Bai, 2015). Caliendo and Kopeinig (2008) state that “the most straightforward matching estimator is nearest neighbor matching”. Other options are kernel matching and radius matching (Heckman et al, 1997; Deheija & Wahba, 2002). This thesis uses these three different methods of propensity score matching to compare results.

Rosenbaum and Rubin (1985) describe nearest neighbor matching as a matching method that “matches each unit in the intervention group with a unit in the control group with the closest absolute distance between their propensity scores”. The intervention group will consist of the hospitals with the largest surgeon teams and the control group will consist of the hospitals with smallest surgeon teams. The variable “surgeon team size” is a binary variable, which means that a hospital either has a small surgeon team size or a large surgeon team size. This variable is described in more detail in the “variables” section. However, Smith and Todd (2005) state that “nearest neighbor matching results in many bad matches, in the sense that many participants get matches to non-participants with very different propensity scores”. I therefore chose to impose a tolerance level

on the maximum propensity score distance, which is defined as a caliper (Cochran & Rubin, 1973). The treated units are matched with its closest control observation within the chosen caliper. Imposing a caliper avoids bad matches and hence the quality of the matches increases. However, there is a risk of oversampling when using this method. I discuss the phenomenon of oversampling later in this part of the thesis.

The second propensity score matching model that I used is kernel matching. Kernel matching is previously described in studies by Heckman et al. (1997) and Heckman et al. (1998), and is defined later by Smith and Todd (2005) as "a nonparametric matching estimator that construct a match for each program participant using a kernel weighted average over multiple persons in the comparison group". Smith and Todd (2005) also state that "weights depend on the distance between each individual from the control group and the observation for which the counterfactual outcome is estimated". Since studies by Heckman et al. (1997) and Heckman and Lozano (2004) recommend a bandwidth parameter of 0.06 and given that this choice did not lead to biased results, I chose for a bandwidth parameter of 0.06. Caliendo and Kopeinig (2008) state in their section about kernel and local linear matching that "a major advantage of these approaches is the lower variance which is achieved because more information is used" and that "a drawback of these methods is that possibly observations are used that are bad matches", which can be a threat to the validity of the results. To correct for these bad matches, I chose to include two options in the matching process and calculate multiple balancing measures. I will elaborate on these choices in the next sections.

Furthermore, the options "with replacement" and "two matches per observation" are selected in the propensity score matching analysis. The option "with replacement" means that a matched observation can be used again to match another observation than its match. Smith and Todd (2005) describe that "allowing replacement results in an increased average matching quality, but also increases the estimator's variance". Caliendo and Kopeinig (2008) state that this option "is of particular interest with data where the propensity score distribution is very different in the treatment and control group", resulting in bad matches. Therefore, this option fits my data well. The drawback to matching with replacement is that the variance will be higher because fewer observations are being used for the implicit comparison group (Bryson et al., 2002). The choice for two matches (with a uniform weight) per observation is made to exclude one-to-one matches that happen to be very good or bad matches by coincidence. In the literature, this is defined as "oversampling". Smith and Todd (2005) summarize this option as follows: "this option trades reduced variance (resulting from using more information to construct the counterfactual for each participant) for increased bias (resulting from using, on average, poorer matches)".

After the implementation of the matching method, the quality of the matching process and the outcomes are evaluated. I calculated the following measures which indicate the extent of balancing of the covariates between the intervention group and the control group and can therefore be used in the search for a suitable matching method and a suitable combination of dependent and independent variables that achieve good balancing.

First, the measures that indicate the variance of the results and the extent to which the results are biased are discussed. I calculated Rubin's B and Rubin's R. Rubin's B is defined by Rubin (2001) as "the absolute standardized difference of the means of the linear index of the propensity score in the intervention group and the control group". This measure represents the extent to which the results of the PSM models are biased. Rubin's R is defined by Rubin (2001) as "the ratio of treatment variance to control variance of the propensity score index". Rubin (2001) states that this variance ratio should equal 1 if there is a perfect balance between the covariates in the intervention group and the covariates in the control group. These two measures of balance across treatment and control group indicate if the results that are obtained from the PSM models are efficient and if the observed bias of the results is minimal (Rubin, 2001). Since Rubin (2001) recommends that "B be less than 25

and that R be between 0.5 and 2 for the samples to be considered sufficiently balanced", I chose these limits as well. This means that before I interpret the results of the PSM models, I first need to check if the Rubin's B and Rubin's R values of the models score between the given limits. If they do, their results can be interpreted as efficient and with minimal bias (Rubin, 2001).

Secondly, the independent t-tests for equality of means in the two samples are discussed. The website of UCLA (n.d.) states that "this t-test assumes that variances for the two samples are the same" and that "this t-test is designed to compare means of the same variable between two groups". The website of Laerd Statistics (n.d.) describes the usage of t-tests as follows: "It can be used to determine whether the mean of a dependent variable is the same in two groups. More specifically, the independent t-test determines whether the mean difference between two groups is statistically significantly different to zero". In terms of the variables that are used in this thesis, the t-test determines whether the mean difference in EQ-5D index score between hospitals with a large surgeon team size and hospitals with a small surgeon team size is statistically significantly different to zero.

3.2 Data and variable construction

3.2.1 Data collection

Data was obtained from the Dutch National health care Institute (Zorginstituut Nederland, 2021). This dataset is composed by the Dutch Orthopedic Association, The Dutch Hospital Association, The Dutch Health Insurer Association and The Dutch Patients Federation. The dataset consists of multiple sets of indicators that give quality information on 44 different medical specialist care procedures for 90 different health care organizations in the Netherlands in 2019. For this thesis, the indicator sets "Hip prosthesis" and "Knee prosthesis" will be used, that are provided alongside the dataset.

The information in the dataset is provided by the Dutch Hospital Data Foundation (Stichting DHD) and consists of information on variables such as the number of surgeons performing THP or TKP in a hospital in a given year, and the overall number of performed procedures per organization. However, since the baseline characteristics or covariates of a hospital are used to compute a propensity score, additional data collection is required to gather more information on these characteristics of these organizations. Among others, I gathered information on the number of beds, the type of hospital and the number of TKPs or THP the hospital performed. A full list of which variables are included can be found in appendix D. I chose these characteristics because based on the theoretical framework and my own expectations, they are predicted to account for most of the differences among hospitals with a large surgeon team size and a small surgeon team size. This information will be gathered from online databases from www.volksgezondheidenzorg.info, www.jaarverantwoordingzorg.nl, www.dhd.nl and www.cbs.nl. The gathered additional information is then merged with the original dataset.

Most of the hospital characteristics are gathered from the previously mentioned databases and then merged with the original database on the basis of the chamber of commerce number of the hospitals and independent treatment centres. The OOR region (Onderwijs- en opleidingsregio) of a hospital is the only variable that is not retrieved from a database. This covariate is created manually in Stata by looking up the geographical location of the hospitals. I elaborate more on this variable in the section "Independent variables".

3.2.2 The sample

I analyze all general, categorical and university hospitals in the Netherlands for which quality data are available in the previously mentioned databases. Health care organizations with limited or no availability to background information or information about the quality of TKP and THP procedures

are excluded from the analysis. This allows me to avoid the problem of variables with incomplete information, because including these observations could be a threat to the validity and reliability of the analysis due to their missing data. 18 Organizations were excluded from the analysis because either their quality data was incomplete, or I had access to limited or no background information on these organizations. This exclusion is a threat to the validity and reliability of the results. However, there is a pattern in the observations that are excluded; they are all independent treatment centres. The reason for this pattern could be that characteristics of independent treatment centres are harder to find or access due to transparency reasons. No hospitals were excluded from the analysis and this resulted in a remaining 72 hospitals in both samples.

3.2.3 Dependent variable

The dependent variable of this thesis is "surgeon team size". Surgeon team size is defined as the amount orthopedic surgeons that places a THP or a TKP. This variable is measured by the indicator that measures the number of orthopedic surgeons that places a primary THP or primary TKP (see Appendix A and Appendix C for a more detailed description). In the dataset, the surgeon team size is a binary variable. The hospitals with a surgeon team size of more than six surgeons are labelled as hospitals with a large surgeon team size (the intervention group). The hospitals with six or less surgeons in their surgeon team are labelled as hospitals with a small surgeon team size, creating the control group. My choice for surgeon teams with six surgeons as large surgeon teams is based on the distribution of surgeon team size in the dataset. When analyzing the surgeon team size distribution in more detail, I found that the 75th percentile of the population had a surgeon team size of six surgeons. I therefore chose to label hospitals with a surgeon team size of more than six surgeons as hospitals with a large surgeon team size, since I only want the largest surgeon teams in the intervention group and the majority of the hospitals in the sample had a surgeon team size of six or less surgeons (75% for TKP surgeon teams and 77.78% for THP surgeon teams). The hospitals are labelled per procedure through dummy variables that indicate whether an observation has a large or small surgeon team size.

3.2.4 Independent variables

For the matching process, I gathered information on the characteristics of hospitals in the sample. These are characteristics such as the profitability of a hospital, the number of beds per hospital at the end of the reporting year, or the number of procedures per surgeon per year. I tried to find information on the determinants of quality of care that are affected by surgeon team size, as discussed in the theoretical framework. For example, information on the coordination of teams or information on the communication of teams. However, this kind of information was hard to find. I therefore only included the surgery volume in the selection of the independent variables, because this was the only type of information where I found (usable) data on. A complete list of the hospital characteristics can be found in appendix D.

During the analysis I encountered problems with the validity and reliability of the results when I included all of the independent variables that are listed in appendix D in the analysis. I therefore made a trade-off between the inclusion of more variables but more biased results and the inclusion of less variables and less biased results. This resulted in a selection of four independent variables that are expected to affect the surgeon team size of a hospital. I included the staff-to-bed ratio, the number of procedures per surgeon per year, the OOR region and the type of hospital. In the following sections I give more information on these independent variables and I discuss the choices I made regarding these variables and why I included them in the analysis.

I first discuss the staff-to-bed ratio. The staff-to-bed ratio is a variable that contains two expected predictors of surgeon team size: the number of beds per hospital at the end of the reporting year and the number of full-time equivalents (FTE's) on December 31 of the reporting year. In the analysis, I wanted to include information on hospital size in terms of beds and personnel. However, the inclusion of these two variables resulted in biased results, because the means for these variables between the intervention and control group differed significantly (p=0.05). To tackle this problem, I created the variable "staff-to-bed ratio". This variable takes the hospital size (in terms of beds and total personnel) into account, because it is defined as the number of beds divided by the number of full-time equivalents. The means of the staff-to-bed ratio between the intervention and control group did not differ significantly (p=0.05). Therefore, the staff-to-bed ratio is used to represent hospital size in terms of the number of beds and personnel in the analysis.

The next independent variable is the number of procedures per surgeon per year. Since there is literature on beneficial effects of an increased volume of procedures on (patient-reported) outcomes (Mesman et al., 2017; Katz et al., 2001), I wanted to include information on the volume of procedures in the analysis. The inclusion of the number of procedures per hospital per year in the analysis resulted in biased results, because the means of this variable between the intervention and control group differed significantly (p=0.05) as well. Therefore, I created a variable that takes hospital size (in terms of the surgeon team size) into account. This is the variable "the number of TKP/THP procedures per surgeon per hospital". This variable is defined as the number of TKP or THP procedures per hospital divided by the number of surgeons per hospital. The mean number of procedures per surgeon per hospital did not differ significantly (p=0.05). Therefore, I included this variable in the analysis to represent the volume-outcome effect.

Furthermore, I describe the OOR regions. I included this variable because I want to explore whether or not the geographical location of the hospitals affects the surgeon team size of a hospital. OOR regions are educational networks in the Netherlands in which the academic hospital of the region works together with general hospitals and other educational organizations in the region. These networks are founded to improve quality, continuity and efficiency of care. However, not every hospital or health care organization in the region is affiliated with these networks. So, for the variable "OOR region" I assumed that hospitals can be included in the regional educational network based on their geographical location, whilst OOR regions in the Netherlands only consist of hospitals that are actually affiliated with the network. This resulted in OOR regions that consist of hospitals that are in fact in the network and hospitals which are located in the same geographic area as the OOR regions (but who are not included in the original OOR regions). Therefore, these OOR regions serve as geographical regions in which the hospitals are located. I created a dummy variable for each region to label hospitals with a certain OOR region. I do not use all dummy categories in the regression. Doing so would give the regression redundant information, result in multicollinearity. In the analysis, I used the OOR region of Zuid West Nederland as reference category because together with the OOR region of Amsterdam Medical Center, this is the OOR region that includes the most observations. Therefore, I am able to see the effects of the more 'uncommon' OOR regions on surgeon team size and quality of care.

Finally, I will discuss the type of hospital. Since the surgeon team size is hypothesized to be affected by factors such as the academic status of a hospital and whether or not a hospital is a specialized hospital, I included the type of hospital as independent variable in the analysis. I created a dummy variable that represented general hospitals and a dummy variable that represented hospitals with an academic status and specialized hospitals. In the analysis, I chose the dummy variable for general hospitals as reference category because this dummy variable included more observations than the dummy variable of specialized hospitals. Therefore, I am able to see the effects of more uncommon hospital types on surgeon team size and quality of care. I chose to merge the academic hospitals and specialized hospitals together because during the analysis, I found that a dummy variable for

academic hospitals predicted failure perfectly and STATA would omit this variable. However, I did want to include information on the academic status of hospitals in the analysis and therefore I merged these two types of hospitals into one dummy variable.

3.2.5 Outcome variable

The difference in patient-reported quality of care after primary hip- and knee replacements is used as an indicator of health care quality for this thesis. In terms of quality of care, modern medicine is shifting to 'personalized medicine'. The focus is not purely on medical treatment anymore and concepts as patient-centeredness are upcoming concepts in orthopedic care (Nederlands Orthopaedisch Vereniging, 2020). This thesis aims to contribute to this shift by taking the patient experiences into account in the measurement of quality of care. Therefore, PROMs (Patient-reported Outcome Measures) are used in the dependent variable. These PROMs measure health outcomes from the perspective of the patient and therefore give insights into the added value of certain forms of care for the patient (Nederlandse Orthopaedische Vereniging, 2020). To do that, PROMs measure the experienced outcomes of the received care, such as pain, functioning and quality of life of patients (NIVEL et al., 2018). The type of PROMs that could be used to represent the outcome variable in the analysis are the EuroQol 5D (EQ-5D), the Numeric (pain) rating scale (NRS-pain), the Hip disability and Osteoarthritis Outcome Score-Physical function Short form (HOOS-PS) and the Knee Injury and Osteoarthritis Outcome Score short form (KOOS-PS). Due to time limitations, I only included the EQ-5D in the analysis. My choice for the EQ-5D index as outcome variable originates from the fact that it is the most widely used generic multi-attribute utility instrument in the world (Jiang et al., 2021). Furthermore, the EQ-5D index takes 5 dimensions of health into account to construct a weighted health index. This makes the EQ-5D a representative instrument for patient-reported quality of care. However, the EQ-5D is based on a visual analogue scale (VAS) and according to Yates et al. (2018), a drawback for this instrument is the end-of-scale bias in which "respondents are less likely to use the extreme ends of the scale to assess their health status". More information on the available PROMS is gathered from www.meetinstrumentenzorg.nl and provided in appendix E.

The outcome variables are operationalized as the difference in PROM-score between the preoperative situation and 12 months after the procedure on the basis of prospective measurement of patients with OA who have undergone a THP or a TKP. The choice for the time path of 12 months after the procedure is influenced by a study of Friebel et al. (2017). Friebel and colleagues mentioned in their article that six months is the earliest time for assessment of benefits in PROMs post-surgery. Therefore, the time path of 12 months after surgery is chosen. A more detailed description of this outcome variable can be found in Appendix B and Appendix C.

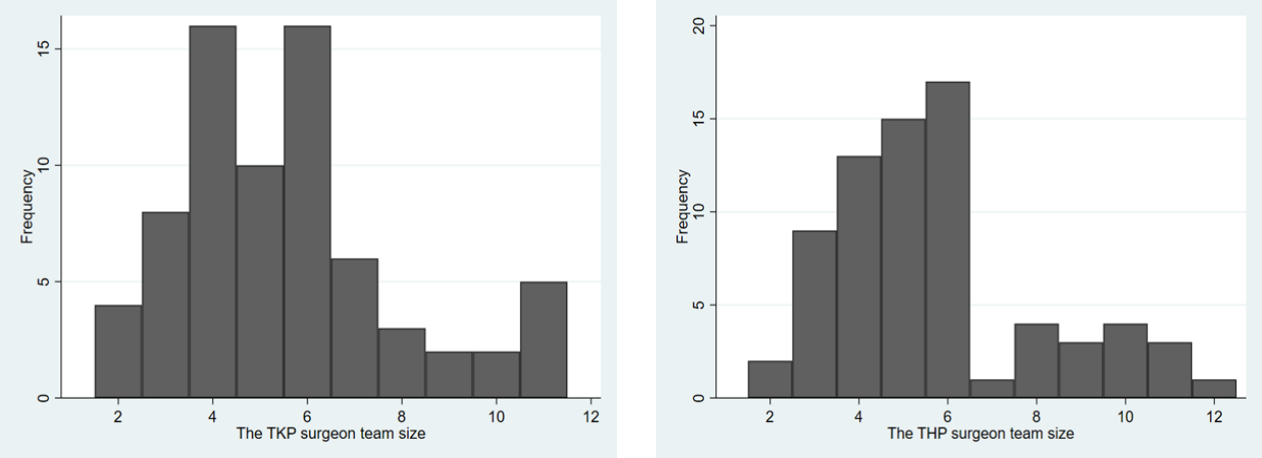
4. Results

4.1 Characteristics of the health care organizations

First, I report the surgeon team sizes of the organizations. To gain insights in the surgeon team size of the hospitals in the sample, I conducted descriptive statistics in STATA. The results of this analysis will be reported in the next section.

The mean TKP surgeon team size for the total sample is 5.57 with a standard deviation of 2.34, while the mean THP surgeon team size for the total sample is 5.71 with a standard deviation of 2.35. The smallest surgeon team size is two surgeons for both procedures and the largest surgeon team size is 11 for TKPs and 12 for THP. Figure 2 represents distributions of the TKP surgeon team size and THP surgeon team size among the sample. As displayed, the surgeon team size is not equally distributed among procedures. The TKP surgeon team size distribution tends more to a normal distribution than the THP surgeon team size, but in this thesis I assume that both procedures have a normal distribution.

Figure 2: histogram with the frequency of the TKP surgeon team size



Note: The left histogram displays the distribution of the TKP surgeon team size among the hospitals and the right histogram displays the distribution of the THP surgeon team size among the hospitals

To gain more insights in the characteristics of hospitals with a small surgeon team size and hospitals with a large surgeon team size, I conducted descriptive statistics in STATA. During this analysis, I used the dummy variable that serves as indicator for hospitals with a large surgeon team size as grouping variable. The results are displayed in Table 1A and table 1B. In the next section I will highlight the differences in the mean values for the covariates between large surgeon team organizations and small surgeon team concerns per procedure.

For instance, concerns with a small surgeon team size have an average surgeon team size of 4.48 surgeons with a standard deviation of 1.27 for TKP and an average surgeon team size of 4.64 surgeons with a standard deviation of 1.18 for THP. Large surgeon team size concerns have an average team size of 8.83 surgeons with a standard deviation of 1.69 for TKP and an average team size of 9.44 surgeons with a standard deviation of 1.41 for THP. Concerns with a small surgeon team size do fewer mean procedures per surgeon per year than large surgeon team size concerns. Large surgeon team concerns also have more mean beds than small surgeon team concerns and have a less

mean total personnel costs. Furthermore, large surgeon team concerns have less mean FTEs employed and a lower staff to bed ratio than small surgeon team size concerns.

Table 1A: descriptive statistics of large surgeon team organizations that perform a TKP compared to small surgeon team organizations that perform a TKP

Hospitals with a small surgeon team size

Variable	N	Mean	Std.Err.	Min	Max
Number of surgeons.	54	4.481	1.270	2	6
Mean number of TKPs per surgeon	54	69.827	28.485	13.667	178.600
Number of beds	54	482.444	267.175	50	1125
Number of full time equivalents (FTE's)	54	3884.453	3281.727	970	13746
Mean total personnel costs	54	2.09e+08	2.15e+08	3.22e+07	9.89e+08
Staff-to-bed ratio	54	7.825	3.822	3.686	22.201

Hospitals with a large surgeon team size

Variable	N	Mean	Std.Err.	Min	Max
Number of surgeons.	18	8.833	1.689	7	11
Mean number of TKPs per surgeon	18	67.637	16.619	44.900	107.625
Number of beds	18	644.667	241.670	167	1103
Number of full time equivalents (FTE's)	18	4192.833	1952.517	1555	9661
Mean total personnel costs	18	2.20e+08	9.19e+07	9.13e+07	4.33e+08
Staff-to-bed ratio	54	6.549	1.310	4.651	9.858

Table 1B: descriptive statistics of large surgeon team organizations that perform a THP compared to small surgeon team organizations that perform a THP

Hospitals with a small surgeon team size

Variable	N	Mean	Std.Err.	Min	Max
Number of surgeons.	56	4.643	1.182	2	6
Mean number of THPs per surgeon	56	70.073	27.934	13.667	178.600
Number of beds	56	479.768	262.338	50	1125
Number of full time equivalents (FTE's)	56	3857.651	3224.644	970	13746
Mean total personnel costs	56	2.08e+08	2.11e+08	3.22e+07	9.89e+08
Staff-to-bed ratio	56	7.815	3.748	3.686	22.201

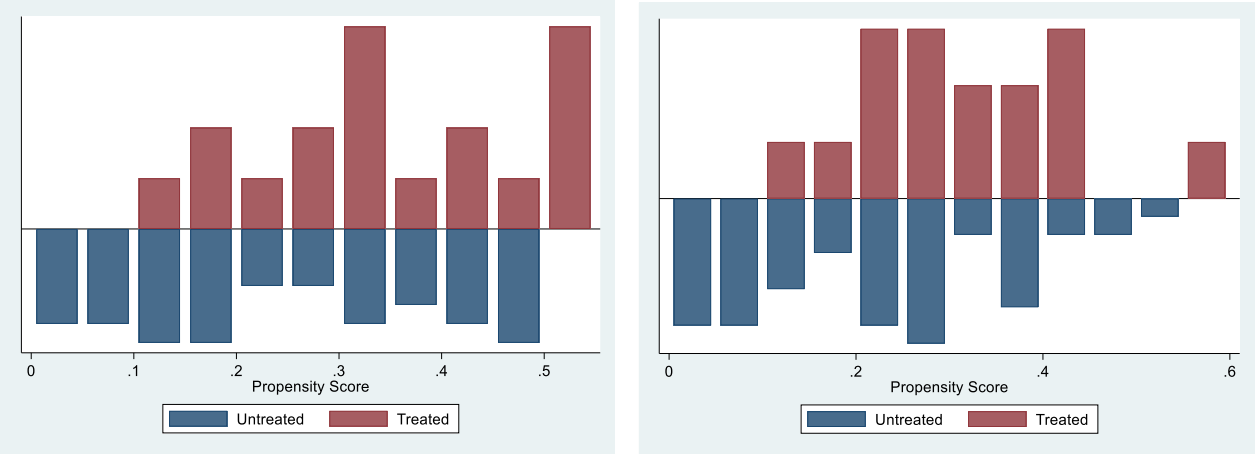
Hospitals with a large surgeon team size

Variable	N	Mean	Std.Err.	Min	Max
Number of surgeons.	16	9.438	1.413	7	12
Mean number of THPs per surgeon	16	66.502	17.676	44.89	107.625
Number of beds	16	674.313	241.560	167	1103
Number of full time equivalents (FTE's)	16	4325.188	2037.335	1555	9661
Mean total personnel costs	16	2.25e+08	9.69e+07	9.13e+07	4.33e+08
Staff-to-bed ratio	16	6.422	1.357	4.651	9.858

4.2 Propensity score matching results

Secondly, I report the propensity score matching results, starting with the quality of the matches. The quality of the matches is assessed by examining the distribution of the propensity scores in the analysis. Figure 3 contains histograms that display density distributions of the propensity scores. The treated group contains hospitals with a large surgeon team size and the untreated group contains hospitals with a small surgeon team size. The figure displays that the propensity scores are distributed similarly between the two procedures. Both distributions lack treated observations and control observations in certain regions. For instance, there are no treated observations in the region [0;0.1] for both distributions and in the THP distribution there are also no treated variables in the [0.45;0.55] region. Furthermore, I do not find any control observations for the treatment observations with a highest propensity scores in both distributions. These 'missings' make the estimation of treatment effects in these regions questionable, given that no observations are excluded based on this criterion (Caliendo & Kopeinig, 2008). I therefore interpret the results of the PSM models with caution.

Figure 3: Propensity score distribution



Note: The left histogram displays the propensity score distribution for TKP procedures and the right histogram displays the propensity score distribution for THP procedures.

To see what the effect of surgeon team size is on patient-reported quality of care per procedure, I have run a nearest neighbor matching model and a kernel matching model on the variables "staff-to-bed ratio", "number of TKP procedures per surgeon", "type of hospital" and "geographical OOR region". The extent to which these covariates are balanced between the intervention group and the control group is indicated with the measures Rubin's B and Rubin's R. Rubin's B indicate the extent to which the results are biased, and Rubin's R represents the variance ratio. Rubin (2001) recommends that "B be less than 25 and that R be between 0.5 and 2 for the samples to be considered sufficiently balanced". If we look at tables 2A and 3A, we see that no model exceeds the thresholds of 0.5 and 2 for Rubin's R and the nearest neighbor matching model for THP exceeds the threshold of 25 for Rubin's B. Therefore, I can conclude that the results in the nearest neighbor matching model for THP are biased and should be interpreted with caution. I report the results in the following tables.

Table 2A: Matching results for TKP

Kernel matching

	Mean			t-test	p-value
Variable	**Large surgeon team size (n=18)**	**Small surgeon team size (n=48)**	**Difference**	**t**	**p>t**
Staff to bed ratio	6.549	6.673	-0.124	-0.210	0.837
Number of TKPs per surgeon	47.592	48.281	-0.689	-0.130	0.896
Indicator for OOR LUMC	0	0	.	.	.
Indicator for OOR Noord- en Oost Nederland	0.222	0.245	-0.023	-0.150	0.878
Indicator for OOR Oost Nederland	0.111	0.132	-0.021	-0.180	0.855
Indicator for OOR UMC	0.167	0.125	0.042	0.340	0.733
Indicator for OOR Zuid Oost Nederland	0.111	0.134	-0.023	-0.210	0.838
Indicator for OOR AMC	0.278	0.229	0.049	0.320	0.747
Indicator for a specialized or academic hospital	0.111	0.132	-0.021	-0.180	0.856

Notes: B=15.5, R=1.20. Each indicator variable represents a dummy variable. Six variables in the control group are dropped, because the indicator variable for the OOR region Leiden University Medical Centre predicts failure perfectly.

Nearest neighbor matching

	Mean			t-test	p- value
Variable	**Large surgeon team size (n=18)**	**Small surgeon team size (n=48)**	**Difference**	**t**	**p>t**
Staff to bed ratio	6.549	6.626	-0.077	-0.140	0.891
Number of TKPs per surgeon	47.592	47.826	-0.234	-0.040	0.966
Indicator for OOR LUMC	0	0	.	.	.
Indicator for OOR Noord- en Oost Nederland	0.222	0.167	0.055	0.410	0.684
Indicator for OOR Oost Nederland	0.111	0.167	-0.057	-0.470	0.641
Indicator for OOR UMC	0.167	0.139	0.028	0.230	0.823
Indicator for OOR Zuid Oost Nederland	0.111	0.139	-0.028	-0.250	0.808
Indicator for OOR AMC	0.278	0.278	0	0.000	1.000
Indicator for a specialized or academic hospital	0.111	0.139	-0.028	-0.250	0.808

Notes: B=23.3, R=1.87. Each indicator variable represents a dummy variable. Six variables in the control group are dropped, because the indicator variable for the OOR region Leiden University Medical Centre predicts failure perfectly.

Table 2B: Treatment effect on the treated estimation results TKPs

Kernel matching

Difference in EQ-5D index score 12 months after surgery	Large surgeon team size (n=18)	Small surgeon team size (n=48)	Difference	Std.Err.	T-stat
Unmatched	0.194	0.192	0.002	0.022	0.070
ATT	0.194	0.187	0.007	0.021	0.350

Note: S.E. does not take into account that the propensity score is estimated.

Nearest neighbor matching

Difference in EQ-5D index score 12 months after surgery	Large surgeon team size (n=18)	Small surgeon team size (n=48)	Difference	Std.Err.	T-stat
Unmatched	0.194	0.192	0.002	0.022	0.070
ATT	0.194	0.184	0.010	0.023	0.430

Note: S.E. does not take into account that the propensity score is estimated.

Table 2C: Probit regression results TKPs

Kernel matching

Indicator for hospitals with a large surgeon team size	Coef.	Std.Err.	z	p-value	[95%Conf.	Interval]
Staff-to-bed ratio	-0.115	0.091	-1.270	0.206	-0.294	0.063
Number of TKPs per surgeon	-0.021	0.011	-1.860	0.063*	-0.042	0.001
Indicator for OOR LUMC	0	(omitted)				
Indicator for OOR Noord- en Oost Nederland	0.510	0.581	0.880	0.380	-0.630	1.649
Indicator for OOR Oost Nederland	0.700	0.700	1.000	0.317	-0.671	2.072
Indicator for OOR UMC	0.514	0.652	0.790	0.430	-0.764	1.792
Indicator for OOR Zuid Oost Nederland	0.413	0.670	0.620	0.538	-0.901	1.727
Indicator for OOR AMC	0.544	0.597	0.910	0.363	-0.627	1.714
Indicator for a specialized or academic hospital	-0.414	0.575	-0.720	0.471	-1.540	0.712

*Note: Six variables in the control group are dropped, because the indicator variable for the OOR region Leiden University Medical Centre predicts failure perfectly. The variable is therefore omitted in the regression. * indicates the significance at a 0.1% level.*

Nearest neighbor matching

Indicator for hospitals with a large surgeon team size	Coef.	Std.Err.	z	p-value	[95%Conf.	Interval]
Staff-to-bed-ratio	-0.106	0.090	-1.180	0.239	-0.281	0.070
Number of TKPs per surgeon	-0.018	0.011	-1.710	0.088*	-0.039	0.003
Indicator for OOR LUMC	0	(omitted)				
Indicator for OOR Noord- en Oost Nederland	0.498	0.578	0.860	0.389	-0.635	1.632
Indicator for OOR Oost Nederland	0.679	0.696	0.980	0.329	-0.685	2.043
Indicator for OOR UMC	0.235	0.626	0.370	0.708	-0.992	1.462
Indicator for OOR Zuid Oost Nederland	0.408	0.668	0.610	0.541	-0.901	1.717
Indicator for OOR AMC	0.564	0.594	0.950	0.343	-0.601	1.728
Indicator for a specialized or academic hospital	-0.372	0.570	-0.650	0.514	-1.489	0.745

*Note: Six variables in the control group are dropped, because the indicator variable for the OOR region Leiden University Medical Centre predicts failure perfectly. The variable is therefore omitted in the regression. * indicates the significance at a 0.1% level.*

Table 3A: Matching results for THPs

Kernel matching

		Mean		t-test	p-value
Variable	Large surgeon team size (n=16)	Small surgeon team size (n=50)	Difference	t	p>t
Staff to bed ratio	6.422	6.448	-0.026	-0.050	0.962
Number of THPs per surgeon	66.502	68.689	-2.187	-0.310	0.759
Indicator for OOR LUMC	0	0	.	.	.
Indicator for OOR Noord- en Oost Nederland	0.250	0.224	0.026	0.170	0.866
Indicator for OOR Oost Nederland	0.125	0.122	0.003	0.030	0.977
Indicator for OOR UMC	0.188	0.186	0.002	0.010	0.991
Indicator for OOR Zuid Oost Nederland	0.188	0.156	0.032	0.230	0.818
Indicator for OOR AMC	0.188	0.260	-0.072	-0.480	0.637
Indicator for a specialized or academic hospital	0.125	0.117	0.008	0.070	0.947

Notes: B=22.4, R=0.89. Each indicator variable represents a dummy variable. Six variables in the control group are dropped, because the indicator variable for the OOR region Leiden University Medical Centre predicts failure perfectly.

Nearest neighbor matching

	Mean			t-test	p-value
Variable	**Large surgeon team size (n=16)**	**Small surgeon team size (n=50)**	**Difference**	**t**	**p>t**
Staff to bed ratio	6.422	6.008	0.414	0.990	0.332
Number of THPs per surgeon	66.502	71.006	-4.504	-0.640	0.526
Indicator for OOR LUMC	0	0	.	.	.
Indicator for OOR Noord- en Oost Nederland	0.250	0.188	0.062	0.420	0.681
Indicator for OOR Oost Nederland	0.125	0.156	-0.031	-0.250	0.807
Indicator for OOR UMC	0.188	0.219	-0.031	-0.210	0.833
Indicator for OOR Zuid Oost Nederland	0.188	0.125	0.063	0.470	0.640
Indicator for OOR AMC	0.188	0.250	-0.062	-0.420	0.681
Indicator for a specialized or academic hospital	0.125	0.094	0.031	0.270	0.786

Notes: B=47.6, R=0.73. Each indicator variable represents a dummy variable. Six variables in the control group are dropped, because the indicator variable for the OOR region Leiden University Medical Centre predicts failure perfectly.

Table 3B: Treatment effect on the treated estimation result THPs

Kernel matching

Difference in EQ-5D index score 12 months after surgery	**Large surgeon team size (n=16)**	**Small surgeon team size (n=50)**	**Difference**	**Std.Err.**	**T-statistic**
Unmatched	0.269	0.258	0.011	0.025	0.440
ATT	0.269	0.259	0.009	0.025	0.380

Note: S.E. does not take into account that the propensity score is estimated.

Nearest neighbor matching

Difference in EQ-5D index score 12 months after surgery	**Large surgeon team size (n=16)**	**Small surgeon team size (n=50)**	**Difference**	**Std.Err.**	**T-statistic**
Unmatched	0.269	0.253	0.016	0.025	0.630
ATT	0.269	0.255	0.014	0.029	0.490

Note: S.E. does not take into account that the propensity score is estimated.

Table 3C: Probit regression results THPs

Kernel matching

Indicator for hospitals with a large surgeon team size	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
Staff-to-bed ratio	-0.156	0.106	-1.470	0.141	-0.363	0.052
Number of THPs per surgeon	-0.012	0.010	-1.220	0.224	-0.031	0.007
Indicator for OOR LUMC	0	(omitted)				
Indicator for OOR Noord- en Oost Nederland	1.056	0.677	1.560	0.119	-0.270	2.382
Indicator for OOR Oost Nederland	1.032	0.766	1.350	0.178	-0.469	2.533
Indicator for OOR UMC	1.179	0.733	1.610	0.108	-0.259	2.616
Indicator for OOR Zuid Oost Nederland	1.221	0.724	1.690	0.092*	-0.199	2.641
Indicator for OOR AMC	0.654	0.677	0.970	0.333	-0.672	1.981
Indicator for a specialized or academic hospital	-0.274	0.591	-0.460	0.643	-1.433	0.885

*Note: Six variables in the control group are dropped, because the indicator variable for the OOR region Leiden University Medical Centre predicts failure perfectly. The variable is therefore omitted in the regression. * indicates the significance at a 0.1% level.*

Nearest neighbor matching

Indicator for hospitals with a large surgeon team size	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
Staff-to-bed ratio	-0.145	0.105	-1.380	0.169	-0.352	0.062
Number of THPs per surgeon	-0.008	0.009	-0.890	0.374	-0.026	0.010
Indicator for OOR LUMC	0	(omitted)				
Indicator for OOR Noord- en Oost Nederland	1.006	0.668	1.500	0.132	-0.304	2.316
Indicator for OOR Oost Nederland	1.022	0.760	1.340	0.179	-0.468	2.512
Indicator for OOR UMC	0.848	0.692	1.230	0.220	-0.507	2.204
Indicator for OOR Zuid Oost	1.198	0.718	1.670	0.095*	-0.208	2.604
Indicator for OOR AMC	0.648	0.671	0.970	0.334	-0.667	1.962
Indicator for a specialized or academic hospital	-0.207	0.590	-0.350	0.726	-1.362	0.949

*Note: Six variables in the control group are dropped, because the indicator variable for the OOR region Leiden University Medical Centre predicts failure perfectly. The variable is therefore omitted in the regression. * indicates the significance at a 0.1% level.*

The results of the propensity score matching analysis are shown in all components of table 2 and table 3, where table 2 represents the results for the TKP procedures and table 3 represents the results for THP procedures. One of the first things to notice is that for all PSM models, the indicator for the OOR region Leiden University Medical Centre (LUMC) is dropped. This variable predicts failure perfectly, because none of the hospitals in the OOR region LUMC has a large surgeon team size. This has resulted in the exclusion of six observations in the analysis and a sample where 18 hospitals with a large surgeon team size are compared to 48 hospitals with a small surgeon team size for TKP and 16 hospitals with a large surgeon team size are compared to 50 hospitals with a small surgeon team size. For the matches to be a good match, the differences in the means for the covariates between the control group and the intervention group should be statistically nonsignificant. If I reflect on this criterion by examining Table 2A and 3A, I can conclude that the matching was rather good since none of the differences are statistically significant.

Table 2B and 3B represent the average treatment effects on the treated (ATT) for both procedures. This parameter represents the effect of surgeon team size on the patient-reported quality of care by comparing hospitals with a large surgeon team size and hospitals with a small surgeon team size their average difference in EQ-5D index score 12 months after surgery. The ATT for TKPs is 0.010 in the nearest neighbor matching model (B=23.3, R=1.87) and 0.007 in the kernel matching model (B=15.5, R=1.20). The ATT for THPs is 0.016 in the nearest neighbor matching model (B=47.6, R=0.73) and 0.014 in the kernel matching model (B=22.4, R=0.89). This means that for both procedures and for both PSM models, the difference between the pre-operative EQ-5D index score and the EQ-5D index score 12 months after surgery is larger for hospitals with a large surgeon team size than for hospitals with a small surgeon team size. However, all ATT estimates are not significant on neither the 5% significance level, nor the 10% significance level.

Table 2C and 3C show results from nearest neighbor matching and kernel matching probit regression models that predicts the dependent variable “Indicator for hospitals with a large surgeon team size”., I can conclude from both regression models for TKPs that the number of TKPs per surgeon per year is negatively and significantly related with the indicator for hospitals with a large surgeon team size at the 10% significance level. Furthermore, I can conclude that both PSM models for THPs, the indicator for OOR region Zuid Oost is positively and significantly related with the indicator for hospitals with a large surgeon team size at the 10% significance level (p=0.095 and p=0.092). However, the coefficient for the indicator for OOR region Zuid Oost in the nearest neighbor matching model must be interpreted with caution since this model’s Rubin’s B value equals 47.6, which exceeds the threshold of 25. The other variables in the models for both procedures do not seem to predict the indicator for hospitals with a large surgeon team size significantly.

5. Discussion and conclusion

Now that I have reported the outcomes of the probit regression and comparison of means that resulted from the propensity score matching analysis, I put these results into context of academic literature. I provided a theoretical framework in which I place this thesis into previous studies regarding the effects of surgeon team size on the quality of care. Studies that have been done in the past focus mainly on the effects of surgical team size on the quality of care, not on the effects of surgeon team size. Nevertheless, these studies provide good lessons and starting points for studies like this thesis, which examines the effects of surgeon team size. These previous studies conclude not only that surgical team size is affected by factors such as organizational and policy context but also that surgical team size affects team performance and therefore quality of care. Furthermore, the effects of surgery volume are discussed in the theoretical framework. Various studies have assessed whether surgery volume affects patient outcome and concluded that there might be a positive association between surgery volume and quality of care. Based on these findings I drafted the two main hypotheses of this thesis, which I test with the help of the results of the PSM analysis in the following sections.

My results show a positive ATT for hospitals with a large surgeon team size on the difference in EQ-5D index score 12 months after surgery. However, these ATT estimates are not considered to be statistically significant. These findings confirm my first hypothesis, in which I state that there is relationship between a hospital's surgeon team size and the hospital's patient-reported quality of care. While the results show a relationship between these two variables, these results lack significance. I am therefore unable to confirm my first hypothesis and conclude that there is a significant relationship between surgeon team size and patient-reported quality of care.

Furthermore, the results show two significant predictors for the indicator of hospitals with a large surgeon team size in the regression. These predictors are surgery volume for TKPs and the indicator for OOR region LUMC for THPs. My second hypothesis suggests that hospitals with a higher surgery volume per surgeon are more likely to be a hospital with a large surgeon team size, while the regression coefficients of surgery volume for TKPs show that hospitals with a higher surgery volume per surgeon are more likely to be a hospital with a large surgeon team size.

The fact that my results differ could be due to the study design because a PSM-analysis only matches on observed information. It may therefore encounter bias from unobserved effects or differences between the intervention and control group. Although a sensitivity analysis shows that most of the results were not driven by differences in case-mix and that the samples are considered sufficiently balanced, there is a possibility that unobserved effects can cause bias in the PSM model estimates. Another reason for these results could be the rather small sample sizes for both groups. This is caused by the small number of hospitals in the Netherlands on which I had information about both the quality of care and hospital characteristics. A large sample size is important for a successful and reliable propensity score matching analysis, since many samples which do not match with the chosen background characteristics are discarded or not used in the analysis. This resulted in a small number of remaining observations and bad matches. This is a threat to the reliability and validity of the results. Therefore, I do not conclude that hospitals with a higher surgery volume are always less likely to have a surgeon team size of 7 or more surgeons. I draw the same conclusion for the relationship between the indicator for OOR region LUMC and surgeon team size.

Now that the results are discussed, I would like to make some suggestions for future research. Given that the results of this thesis are either not significant or need to be interpreted with caution due to bias or bad matches, I am unable to provide any practical recommendations for policy makers or health care organization regarding the effects of surgeon team size on the quality of care. However, I would like to provide insights for future research. In the first place I would like to suggest that future research explores different research methods to look into the effect of surgeon team size on quality of care. I used a propensity score matching analysis, which meant that I could use data from one year (2019) and therefore find a good match between hospitals with a large surgeon team size and hospitals with a small surgeon team size in the same year. The analysis of this thesis therefore acted as an empirical test to find out whether the potential effect of surgeon team size on quality of care can be explored with data from one year. Since the analysis of this thesis resulted in biased and mostly nonsignificant results, I suggest that future research tries to find other research methods that include data on multiple years.

Continuing on the previous recommendation on finding data that includes information on multiple years, future research should also aim at increasing the sample sizes of the analysis. I was only able to include 72 observations in the analysis, of which 6 observations were dropped due to perfect prediction. If I had access to more information on hospital characteristics, I could have included the 18 observations that are currently excluded from the analysis. I therefore suggest that future research tries to find and include more observations in the analysis, so that the matches in the PSM will be better and the validity and reliability of the results are guaranteed. The data collection of these hospital characteristics was a time-consuming process which resulted in my choice for EQ-5D index as only outcome variable due to time limitations. I therefore also suggest that future research tries to expand their analyses with multiple outcome variables such as the procedure specific PROMs “HOOS-PS” and “KOOS-PS”. Finally, I suggest that future research tries to expand their analysis with more covariates to avoid any hidden bias due to latent variables. The inclusion of more covariates will result in more precise estimations of the predictor coefficient.

Finally, I discuss the conclusions of the results. This thesis focused on exploring potential predictors of surgeon team size and look into potential effects between surgeon team size and patient-reported quality of care, where surgeon team size is defined as the number of surgeons per hospital that performs these procedures. To do this, I drafted two hypotheses that incorporate expectations of these predictors and effects based on previous studies. The first hypothesis stated that there is a significant relationship between a hospital’s surgeon team size and the hospital’s patient-reported quality of care. The second and final hypothesis aims at explaining potential differences in quality of care, as hypothesized by the first hypothesis. This second hypothesis stated that there is a significant positive relationship between surgery volume and a hospital’s surgeon team size. By conducting a propensity score matching analysis in STATA, I find that no evidence that supports my first hypothesis. However, for my second hypothesis I find two predictors, in the form of surgery volume and OOR region Leiden University Medical Centre (LUMC), that predict surgeon team size significantly. Unfortunately, I was not able to draw a conclusion regarding my second hypothesis due to biased results and a poor propensity score distribution. However, I am optimistic for results in the future since there is little evidence is available on teamwork in hospital-wide surgeon teams in the current literature. I therefore conclude that future research is needed to gain reliable and valid insights in the effects of surgeon team size on patient-reported quality of care.

Literature

Austin, P. C. (2010). Statistical criteria for selecting the optimal number of untreated subjects matched to each treated subject when using many-to-one matching on the propensity score. *American journal of epidemiology*, 172(9), 1092-1097.

Borrill, C. S., Carletta, J., Carter, A., Dawson, J. F., Garrod, S., Rees, A., ... & West, M. A. (2000). The effectiveness of health care teams in the National Health Service. Birmingham: University of Aston in Birmingham.

Bryson, A., Dorsett, R., & Purdon, S. (2002). The use of propensity score matching in the evaluation of active labour market policies. [PDF]. Consulted from http://eprints.lse.ac.uk/4993/1/The_use_of_propensity_score_matching_in_the_evaluation_of_active_labour_market_policies.pdf

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, 22(1), 31-72.

Cochran, W. G., & Rubin, D. B. (1973). Controlling bias in observational studies: A review. *Sankhyā: The Indian Journal of Statistics, Series A*, 417-446. Consulted from https://www-jstor-org.eur.idm.oclc.org/stable/pdf/25049893.pdf?refreqid=excelsior%3A75714ba35b39e5d348e4762dbf3b13cd

Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and statistics*, *84*(1), 151-161.

Friebel, R., Dharmarajan, K., Krumholz, H. M., & Steventon, A. (2017). Reductions in readmission rates are associated with modest improvements in patient-reported health gains following hip and knee replacement in England. *Medical care*, *55*(9), 834.

Healey, A. N., Undre, S., & Vincent, C. A. (2006). Defining the technical skills of teamwork in surgery. *BMJ Quality & Safety*, 15(4), 231-234.

Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The review of economic studies*, *64*(4), 605-654.

Heckman, J. J., Ichimura, H., Smith, J. A., & Todd, P. E. (1998). *Characterizing selection bias using experimental data*.

Heckman, J., & Navarro-Lozano, S. (2004). Using matching, instrumental variables, and control functions to estimate economic choice models. *Review of Economics and statistics*, *86*(1), 30-57.

Jiang, R., Janssen, M. B., & Pickard, A. S. (2021). US population norms for the EQ-5D-5L and comparison of norms from face-to-face and online samples. *Quality of Life Research*, *30*(3), 803-816.

Katz, J. N., Losina, E., Barrett, J., Phillips, C. B., Mahomed, N. N., Lew, R. A., ... & Baron, J. A. (2001). Association between hospital and surgeon procedure volume and outcomes of total hip replacement in the United States Medicare population. *Journal of Bone and Joint Surgery*, 83(11), 1622-1629.

Kruse, F. M., van Nieuw Amerongen, M. C., Borghans, I., Groenewoud, A. S., Adang, E., & Jeurissen, P. P. T. (2019). Is there a volume-quality relationship within the independent treatment centre sector? a longitudinal analysis. *BMC health services research*, 19(1), 1-13.

Liang, P. J., Rajan, M. V., & Ray, K. (2008). Optimal team size and monitoring in organizations. *The Accounting Review*, 83(3), 789-822.

Makary, M. A., Sexton, J. B., Freischlag, J. A., Holzmueller, C. G., Millman, E. A., Rowen, L., & Pronovost, P. J. (2006). Operating room teamwork among physicians and nurses: teamwork in the eye of the beholder. *Journal of the American College of Surgeons*, 202(5), 746-752.

Mickan, S., & Rodger, S. (2000). The organisational context for teamwork: comparing health care and business literature. *Australian Health Review*, *23*(1), 179-192. doi: 10.1071/ah000179..

Nederlandse Orthopaedische Vereniging. (2020). *PROMs*. Geraadpleegd op 26 maart 2021, van https://www.orthopeden.org/kwaliteit/kwaliteitsbeleid/promsNIVEL, IQ health care, VSOP, Patiëntenfederatie Nederland, Zorginstituut Nederland. (2018) *PROM-wijzer* [Internet]. Consulted from https://www.zorginzicht.nl/ontwikkeltools/prom-toolbox/promwijzer-1.-wat-zijn-proms

Pan, W., & Bai, H. (2015). *Propensity score matching analysis*. Guilford Publications. [PDF]. Consulted from http://people.duke.edu/~wp40/sample_files/chapter%201.pdf

Paris, C. R., Salas, E., & Cannon-Bowers, J. A. (2000). Teamwork in multi-person systems: a review and analysis. *Ergonomics*, 43(8), 1052-1075.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*(1), 33-38.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, 2(3), 169-188.

Salas, E., Wilson, K. A., Murphy, C. E., King, H., & Salisbury, M. (2008). Communicating, coordinating, and cooperating when lives depend on it: tips for teamwork. *The Joint Commission Journal on Quality and Patient Safety*, *34*(6), 333-341.

Schafer, J. L., & Kang, J. (2008). Average causal effects from nonrandomized studies: a practical guide and simulated example. *Psychological methods*, 13(4), 279.

Smith, J. A., & Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators?. *Journal of econometrics*, *125*(1-2), 305-353.

Laerd Statistics (n.d.). *Independent t-test using Stata*. Consulted from https://statistics.laerd.com/stata-tutorials/independent-t-test-using-stata.php

Thompson, B. M., Haidet, P., Borges, N. J., Carchedi, L. R., Roman, B. J., Townsend, M. H., ... & Levine, R. E. (2015). Team cohesiveness, team size and team performance in team-based learning teams. *Medical education*, 49(4), 379-385.

Tilbury-Werkhoven, C. (2018). The value of total hip and knee arthroplasties for patients. [Thesis]. Consulted from www.scholarlypublications.universiteitleiden.nl/access/item%3A2906616/download

UCLA: Statistical Consulting Group. (n.d.). *T-test | Stata Annotated Output.* Consulted from https://stats.idre.ucla.edu/stata/output/t-test/

Vincent, C. (2011). *Patient safety*. John Wiley & Sons. Geraadpleegd op 20 maart 2021, van http://www.siakadkh.com/admin/files_ebook/Patient_Safety__2nd_Edition.pdf

Welch, W. P., Cuellar, A. E., Stearns, S. C., & Bindman, A. B. (2013). Proportion of physicians in large group practices continued to grow in 2009–11. *Health Affairs*, 32(9), 1659-1666.

Yates, H., Adamali, H. I., Maskell, N., Barratt, S., & Sharp, C. (2018). Visual analogue scales for interstitial lung disease: a prospective validation study. *QJM: An International Journal of Medicine*, *111*(8), 531-539.

Zorginstituut Nederland. (2021, 10 maart). *Open data Ziekenhuizen en Zelfstandige Behandelcentra/Medisch-specialistische zorg*. Consulted from https://www.zorginzicht.nl/openbare-data/open-data-ziekenhuizen-en-zelfstandige-behandelcentra---medisch-specialistische-zorg

Appendix A - indicator tables of surgeon team size

Appendix A consists of tables with information about TKP surgery indicators. This information is provided by the indicator guide "Indicatorenset Knieprothese" and also applies for the indicator set of THR surgery, because the indicators are comparable.

Indicatornaam	Aantal orthopedisch chirurgen dat primaire knieprothesen plaatst
Indicatornummer	7b
Operationalisatie	Aantal orthopedisch chirurgen dat primaire knieprothesen plaatst per ziekenhuislocatie
Informatie voor cliënten	Weergave van het aantal orthopedisch chirurgen dat primaire knieprothesen plaatst, per ziekenhuis(locatie). Totale knieprothesechirurgie is een basisvaardigheid voor orthopedisch chirurgen.
Transparantie	verplicht
Type indicator	proces
Relevantie	Deze indicator is vooral geschikt om de overige TKP indicatoren in een context te plaatsen.
Datatype	aantal
Bron (achtergrond) van de indicator	
Rekenregels en definities	
Vraag	Hoeveel orthopedisch chirurgen op uw ziekenhuislocatie plaatsen primaire knieprothesen?
Antwoordopties	één antwoord mogelijk
Definitie	• Primaire knieprothesen: Totale knieprothesen en unicondylaire knieprothesen • Orthopedisch chirurgen: Orthopedisch chirurgen die de operatie zelf uitvoeren of supervisie geven aan AIOS/ANIOS
In-/exclusiecriteria	Het gaat hier alleen om de specialisten die in het hele verslagjaar werkzaam zijn geweest op de betreffende ziekenhuislocatie.
Casemix	Op deze indicator wordt geen casemix correctie toegepast
Databron (registratie)	ZIS (invul)
Meetperiode	Volledige verslagjaar (1 januari tot en met 31 december)
Aanleverfrequentie	één keer per jaar
Aanleverniveau	locatieniveau

Appendix B - indicator table PROMs-score

Appendix B consists of tables with information about TKP surgery indicators. This information is provided by the indicator guide "Indicatorenset Knieprothese" and also applies for the indicator set of THR surgery, because the indicators are comparable.

Indicatornaam	Verschilscore PROMs knie, 12 maanden
Indicatornummer	4d
Operationalisatie	Verschilscore tussen knie PROM preoperatief en 12 maanden na ingreep o.b.v. prospectieve meting van patiënten met artrose bij wie een totale knieprothese wordt geplaatst.
Informatie voor cliënten	Orthopedisch chirurgen willen weten of zij kwalitatief goede zorg leveren. Het perspectief van de patiënt is daarbij van belang (heeft de patiënt minder klachten dan voorheen). Een PROM vragenlijst geeft de mening en waardering weer van de patiënt over zijn of haar eigen gezondheid. Voor patiënten die operatief worden behandeld wordt een PROM-score gemeten die betrekking heeft op het gewricht dat klachten veroorzaakt. Door de patiënt te vragen om voor én na de operatie dezelfde vragenlijst in te vullen kan gekeken worden hoe bepaalde ervaringen en klachten door de operatie veranderen. De behandeling wordt zo behalve met klinische parameters ook geëvalueerd op basis van PROMs. Ook voor de patiënt kan het inzichtelijk maken van zijn of haar voortgang/revalidatie motiverend werken. Er wordt vanuit de NOV gestreefd naar landelijk gebruik van PROMs.
Transparantie	verplicht
Type indicator	uitkomst
Relevantie	Inzichtelijk maken van de verandering op de PROM scores, 12 maanden na het plaatsen van een knieprothese. Voor zowel orthopeed als patiënt relevant.
Datatype	per PROM: Aantal (N) Gemiddelde 95% Betrouwbaarheidsinterval Lower bound 95% Betrouwbaarheidsinterval Upper bound
Bron (achtergrond) van de indicator	NOV PROMs-advies: https://www.orthopeden.org/downloads/775/nov-proms-advies.pdf en
	https://richtlijnendatabase.nl/richtlijn/totale_knieprothese/patient-reported_outcome_measures_bij_tkp.html
Rekenregels en definities	
Vraag	De knie PROM bestaat uit: EQ-5D index score EQ-5D thermometer NRS-pijn rust NRS-pijn activiteit
Definitie	Zorginzicht; wat zijn PROMs? https://www.zorginzicht.nl/kennisbank/Paginas/PROM-wijzer-1-Wat-zijn-PROMs.aspx EQ5D: EQ-5D index score meet de kwaliteit van leven, Een groter verschil staat voor een grotere verandering in kwaliteit van leven. EQ-5D thermometer meet de situatie van de gezondheid. Een groter verschil staat voor een grotere verandering in situatie van de gezondheid. NRS-pijn: meet de pijn tijdens rust en activiteit. Een groter verschil staat voor een grotere verandering in ervaren pijn.
In-/exclusiecriteria	Inclusie: PROMs, benoemd in het NOV PROMs-advies
Casemix	Uitkomst gecorrigeerd voor case mix (Geslacht, Leeftijd, Charnley score, roken, ASA, preoperatieve PROM-score en BMI)
Databron (registratie)	LROI
Meetperiode	Patiënten primair geopereerd in de periode van 1 januari tot en met 31 december voorafgaande aan het verslagjaar.
Aanleverfrequentie	één keer per jaar
Aanleverniveau	locatieniveau

Appendix C – explanation of information per indicator

Appendix C consists of tables with information about TKP surgery indicators. This information is provided by the indicator guide "Indicatorenset Knieprothese" and also applies for the indicator set of THR surgery, because the indicators are comparable.

Operationalisatie	De indicator in één korte zin omschreven. Let op: vermeld duidelijk de eenheid van de indicator in deze zin. Bijvoorbeeld: 'wachttijd in dagen'.
Informatie voor cliënten	Het belang van en de betekenis van de indicator wordt hier kort zonder vaktermen verwoord. Een indicator is een meetbaar onderdeel van de zorg wat iets kan zeggen over de kwaliteit van zorg. In de informatie van cliënten wordt beknopt omschreven wat de indicator betekent en hoe deze geïnterpreteerd moet worden ('lager is beter', 'een instelling moet onder de norm van X scoren').
Transparantie	- verplicht (publicatie in Openbare Database van Zorginstituut Nederland) - vrijwillig (geen openbare publicatie. Doorlevering alleen naar patiëntenorganisaties, zorgverzekeraars en zorgaanbieders)
Type indicator	- Uitkomst - Proces - Structuur
Relevantie	Geef aan waarom de indicator relevant is en voor wie.
Datatype	Het datatype dat moet worden aangeleverd: - tekst (vrije tekst of een keuze uit een lijst in de indicatorgids) - ja/nee - aantal (een geheel getal) - getal - percentage (een getal tussen 0 en 100. Teller/noemer *100)
Bron (achtergrond) van de indicator	Op basis waarvan is de indicator opgesteld? Verwijs naar een richtlijn/standaard, of een internationale indicatorset waarin de indicator ook is opgenomen. Dit vergroot de validiteit van de indicator: zegt deze indicator echt iets over kwaliteit van zorg?
Rekenregels en definities	
Teller(s)	Het getal boven de streep van een breuk. De teller is altijd een deelverzameling van de noemer. Om lange formuleringen te vermijden, is de volledige omschrijving van de deelverzameling niet altijd herhaald in de teller.
Noemer	Het getal onder de streep van een breuk. Nauwkeurige beschrijving van de cliëntenpopulatie. Indien er sprake van een structuurindicator is, dan is noemer niet van toepassing.
Vraag	Wanneer er een vraag wordt gesteld over de organisatie van de zorg (vaak een klantpreferentievraag), dan kan de vraag aan de instelling hier geplaatst worden. De operationalisatie is dan hoe de indicator wordt gepubliceerd. Bijvoorbeeld: Vraag: "Op welke manier kunnen patiënten na de operatie contact opnemen bij vragen?" Operationalisatie: "Aangeboden manieren post-operatief contact"
Antwoordopties	Bij vragen is het belangrijk aan te geven of er slechts één antwoord mogelijk is of meerdere antwoorden mogelijk zijn. Daarnaast moeten de antwoordopties vermeld worden.
Definitie	Indien in de indicator termen worden gebruikt die enige toelichting nodig hebben, zodat betrouwbaar kan worden geregistreerd, dan wordt een definitie gegeven.

In-/exclusiecriteria	Een duidelijke definiëring van de cliëntenpopulatie vertaalt zich uiteindelijk in duidelijke in- en exclusiecriteria. Daarnaast kunnen exclusiecriteria gebruikt worden om vergelijkbaarheid te vergroten. Bijvoorbeeld als bepaalde cliëntengroepen niet gelijk over instellingen zijn verdeeld.
Casemix	Wanneer het relevant is voor een indicator kunnen cliëntkenmerken gebruikt worden voor het corrigeren van de indicatorwaarde. Hier moet worden aangegeven of er een casemixcorrectie plaatsvindt en op welke variabelen.
Databron (registratie)	De te gebruiken bron voor het berekenen van de indicatorwaarde. Bijvoorbeeld: LROI, NKR, DLCA-R, EPD, Zorgkaart Nederland
Norm	Als de indicator een norm kent, wordt deze hier in de indicatorgids vermeld. Ook de bron van de norm wordt vermeld.
Meetperiode	De meetperiode is de periode waarin de metingen worden gedaan. Dit is standaard het hele kalenderjaar (01-01 t/m 31-12), maar hier kan van worden afgeweken. Bij follow-upmetingen moet een expliciete keuze worden gemaakt: een meting drie maanden post-operatief binnen het verslagjaar betekent dat de operatie ook in de laatste drie maanden van het voorgaande jaar kan zijn uitgevoerd. Soms wordt er een peildatum gebruikt in plaats van een meetperiode (vaak bij structuurindicatoren). De peildatum ligt dan vaak op 01-03 van het jaar ná het verslagjaar.
Aanleverfrequentie	De frequentie waarmee de indicatoren aangeleverd worden. Afspraken over de frequentie worden landelijk gemaakt. Dit is momenteel één keer per jaar.
Aanleverniveau	Het niveau waarop de indicatoren worden aangeleverd; in beginsel worden alle indicatoren op locatieniveau aangeleverd. In de uitgangspunten in de indicatorgids wordt voor de hele set aangegeven hoe hier mee om wordt gegaan. Per indicator kunnen specifieke aanwijzingen worden gegeven, bijvoorbeeld voor proces- en uitkomstindicatoren afkomstig uit kwaliteitsregistraties op concernniveau.

Appendix D – list of covariates

Variable	Operationalization
nTKP_per_surgeon and nTHP_per_surgeon	The number of TKP or THP procedures divided by the number of surgeons
type_genhospital	Indicator for general hospitals
type_spechospital	Indicator for university hospitals or hospitals that are specialized in performing a TKP or THP
Staff_to_bed_ratio	The number of full time equivalents on December 31 of reporting year divided by the number of beds at the end of the reporting year
nBed_hospitals	The number of beds at the end of reporting year
total_personnel_costs	The total costs of personnel - amount in Euro's at the end of reporting year
OOR_region	Indicator for the geographical area where the organization is located
result_after_taxes	The result of the organization after taxation - amount in Euro's at the end of reporting year
PersTot_nFTE	The total amount of personnel - The number of full time equivalents (FTE's) on December 31 of the reporting year
Profitability	The degree to which a business or activity yields profit or financial gain. Calculated by the formula "Operating profit for financial income and expenses / balance sheet total"
Liquidity	The ability of a firm to pay its short term obligation for the continuous operation. Calculated by the formula "Current ratio=Current assets / Current liabilities" in the dataset

Appendix E – available types of PROMs

PROM-type	Definition from www.meetinstrumentenzorg.nl
EQ-5D	The EQ-5D is a standardized instrument that scores on five dimensions of health (mobility, self-care, daily activities, pain/discomfort and anxiety/depression). A weighted health index for an individual or population can be derived from this. EuroQol is complementary to other quality of life measuring instruments (such as SF-36). Furthermore, the patient must indicate how he experiences his health status on a scale from 0 to 100. A higher score represents a better health situation (score varies between 0 (worst imaginable health) and 100 (best imaginable health).
HOOS-PS	The HOOS-PS is a questionnaire for evaluating symptoms and limitations in patients with hip complaints. The questionnaire consists of 5 items and is an abbreviated version of the HOOS and is composed of the subcategory activities of daily living (ADL activities) and Sports & Recreation. The short version is about how much effort it took to perform an activity during the past week. Scoring is done using a 5-point Likert scale (0-4), with a higher score indicating more effort.
KOOS-PS	The KOOS-PS is a questionnaire for evaluating symptoms and limitations in patients with knee complaints. The questionnaire consists of 7 items and is an abbreviated version of the KOOS and is from the subcategory ADL activities and Sports & Recreation. The abbreviated version of the effort involved in an activity during the past week. Scoring is done using a 5-point Likert scale (0-4), with a larger score giving more effort.
NRS-pain	The numeric rating scale (NRS) is a non-specific measurement scale, consisting of 11 numbers from 0-10, where 0 means no pain at all and 10 is the most imaginable pain. On the left side is the minimum score, on the right side is the maximum score. The patient should circle the number that best represents the severity of his/her pain the patient has had in the past

	week. Because only a whole point can be assigned to the sensation, the NRS is less sensitive to detecting small changes than the VAS.