Established in 1871

Swiss Medical Weekly

Formerly: Schweizerische Medizinische Wochenschrift

An open access, online journal • www.smw.ch

Original article | Published 18 November 2018 | doi:10.4414/smw.2018.14680
Cite this as: Swiss Med Wkly. 2018;148:w14680

Rationale and methods of an observational study to support the design of a nationwide surgical registry: the MIDAS study

Vach Werner[a], Saxer Franziska[a], Holsgaard-Larsen Anders[b], Overgaard Søren[b], Farin-Glattacker Erik[c], Bless Nicolas[a], Bucher Heiner C.[d], Jakob Marcel[a]

[a] Department of Orthopaedics and Traumatology, University Hospital Basel, Switzerland
[b] Orthopaedic Research Unit, Department of Clinical Research, University of Southern Denmark and Department of Orthopaedics and Traumatology, Odense University, Odense C, Denmark
[c] Sektion Versorgungsforschung und Rehabilitationsforschung, Medical Faculty and University Medical Centre, University of Freiburg, Germany
[d] Basel Institute for Clinical Epidemiology and Biostatistics, University Hospital Basel and Department of Clinical Research, University of Basel, Switzerland

Summary

BACKGROUND: Surgical registries are becoming increasingly popular. In addition, Swiss legislation requires data on therapeutic outcome quality. The Swiss Association of Surgeons (Schweizerische Gesellschaft Chirurgie, SGC-SCC) has already agreed on a first minimum data set. However, in the long run the scope and content of the registry should be evidence-based and not only accepted by professional stakeholders. The MIDAS study aims at providing such evidence for the example population of patients undergoing emergency or elective hip surgery. Five relevant aspects are considered: (1) choice of instruments for assessing health related quality of life (HRQoL); (2) optimal time-point for assessment; (3) use of proxy assessments; (4) choice of pre-surgery risk factors; and (5) assessment of peri- and postoperative variables.

METHODS: MIDAS is a longitudinal observational multicentre study. All patients suffering from a femoral neck fracture or from arthritis of the hip joint with an indication for prosthetic joint replacement surgery will be offered participation. The study is based on a combination of routine data from clinical standard practice with specifically documented data to be reported by the treating clinician and data to be collected in cooperation with the patient – in particular patient-reported outcome measures (PROMs). The latter include the Health Utility Index Mark 3 (HUI3) and Euro-Qol-5D (EQ-5D) as generic instruments, Hip Disability and Osteoarthritis Outcome Score (HOOS) as a disease specific instrument for the assessment of HRQoL, and two performance-based functional tests. Data will be collected at baseline, during hospitalisation/at discharge and at three routine follow-up visits. All patients will be asked to name a person for assessing proxy-perceived HRQoL.

DISCUSSION: To the best of our knowledge, this is the first study explicitly addressing questions about the design of a national surgical registry in an empirical manner. The study aims at providing a scientific base for decisions regarding scope and content of a potential national Swiss surgical registry. We designed a pragmatic study to envision data collection in a national registry with the option of specifying isolated research questions of interest. One focus of the study is the use of PROMs, and we hope that our study and their results will inspire also other surgical registries to take this important step forward.

Trial registration: Registered at the "Deutsches Register Klinischer Studien (DRKS)", the German Clinical Trials Registry, since this registry meets the scope and methodology of the proposed study. Registration no.: DRKS00012991

Keywords: *registry*, *surgical research*, *patient reported outcomes*, *complications*, *design*

Author contributions
WV and FS contributed equally. FS and MJ developed the idea of the clinical study in collaboration with the political stakeholders SGC and ANQ. WV refined the research questions and designed the analytic strategy. AHL, SO and EFG gave input on the choice of instruments. HB supported the development of the study design and positioning the study in the field of outcome research and health-economic evaluations. NB implemented pathways for the inclusion of emergency patients and was substantially involved in the development of the implementation of the trial in daily practice. FS and WV wrote the study protocol and drafted the manuscript. All authors read and approved the final manuscript.

Correspondence:
Werner Vach, PhD, Franziska Saxer, MD, Department of Orthopaedics and Traumatology, University Hospital Basel, Spitalstrasse 21, CH-4031 Basel, Switzerland, werner.vach[at]usb.ch, franziska.saxer[at]usb.ch

Introduction

Background

The first nationwide databases on surgical procedures were established in the 1970s in the Nordic countries, with an increasing number of databases founded in different countries over recent decades [1, 2]. This development is on one hand based on scientific interest, and on the other hand reflects legal requirements in many countries. The role of these registries is expanding in research [3–7], quality control [8] and education [9]. The question of how to build up a successful surgical registry has recently been the topic of a systematic review [10].

Swiss legislation requires professional healthcare providers to collect data on quality control and medical outcome parameters. These data should serve several aims – as also discussed in general in relation to surgical registries [1]:

– to inform surgeons about the outcomes of their individual patients;

Published under the copyright license "Attribution – Non-Commercial – No Derivatives 4.0".
No commercial reuse without permission. See http://emh.ch/en/services/permissions.html.

– to inform clinical departments about the distribution of patient outcomes in specific patient groups;
– to inform patients about hospital-specific expected outcomes, in particular long term outcomes including health-related quality of life (HRQoL);
– to inform epidemiological and health economic studies on injury-related disability;
– to allow in all these activities differences to be taken into account in individual, preoperative risk;
– To allow indication-related and treatment-related information to be taken into account in all these activities.

Accordingly, the data should be generic rather than disease- or discipline-specific to allow comparison between different specialties and, in the long run, also between surgical and nonsurgical treatment approaches. Data should be reliable, that is, validated for different patient collectives with validated questionnaires. Data should furthermore be quickly and easily documented either in a face-to-face situation or as a telephone/written survey to ensure a high return rate and prevent sloppy documentation or an addition to the already relevant administrative burden in healthcare. Finally, data should be meaningful in the sense that they cover relevant constructs and allow measurement of relevant differences in quality of healthcare provision.

Currently in Switzerland, standardised documentation is imperative only for pathologies listed as requiring highly specialised medical care, such as polytrauma management, transplantation or oesophageal resection. In addition, there is an obligatory registry for hip and knee arthroplasty, and standardised documentation in the context of certified oncological centres. For all other surgical procedures, data on complications are currently used for an evaluation of healthcare quality. However, the presence of complications does not necessarily imply the absence of quality or vice versa. In spite of the legislative requirement, quality control documentation is currently not implemented nationwide and does not allow comparison of quality since there is no standardised data collection across different disciplines. Even within one discipline there is no standard or consensus on the responsibility for data collection (e.g., the surgeon performing the intervention vs the surgeon/doctor providing in- or out-patient care or administrative personnel), or on the type of data that should be collected, such as the duration of follow-up, the definition of complications (disease specific vs general) or the grading of complications (e.g., according to the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use Guideline for Good Clinical Practice [ICH-GCP] [11] or the Clavien-Dindo classification [12]).

In this situation, the Swiss Association of Surgeons (Schweizerische Gesellschaft Chirurgie; SGC-SSC) together with the Arbeitsgemeinschaft für Qualitätssicherung in der Chirurgie (AQC) has made an effort to define a minimum dataset (MDS) that could ensure the collection of reliable and meaningful generic information on the quality of medical care. The proposed dataset (table 1) has been approved by the SGC-SSC and its affiliated associations (the stakeholders of general, trauma, visceral, vascular, thoracic and hand surgery). However, it is unclear whether this proposal is fully adequate and sufficient.

There are several issues that are unresolved. The following five questions seem to be of central relevance.

I. Choice of instruments for assessing health related quality of life

Instruments based on patient-reported outcome measures (PROMs) are the standard for HRQoL assessment today. However, there is a plethora of generic and disease-specific instruments. In order to allow epidemiological and health economic studies across different pathologies and disciplines, use of a combination of two generic instruments – the Health Utility Index Mark 3 (HUI3) [19] and Euro-Qol-5D (EQ-5D) [16] – has been recommended [20]. On the other hand, objective functional tests such as walking tests or hand grip measurements are available, and these may provide surrogate variables for quality of life. It is unclear which of these choices is the most appropriate and, in particular, whether the choice of two generic instruments to be applied in all patients instead of a large number of disease specific instruments each to be applied in a specific subgroup can be justified.

II. Time-point of follow-up assessment

Apart from the baseline assessment of HRQoL, it is desirable to have only one follow-up assessment at a specific time-point for all patients, at least within a specific patient population. Patients reach a stable level of their HRQoL only some months after surgery, towards the end of their rehabilitation phase. To cover the full impact of surgical care on patients, it is hence desirable to choose a time-point at which the vast majority of patients have reached this stable level. This time-point is unknown for many patient groups. For administrative reasons and to achieve a high return rate, a time-point close to or shortly after regular contact with the healthcare provider would be preferable.

III. Using proxies for the assessment

Surgery is often performed in elderly patients. In consequence, a substantial fraction of patients with cognitive impairment is to be expected, in particular as cognitive

Table 1: The minimal dataset proposed by the SGC-SSC and AQC.

Hospitalisation data	Hospital reference number Patient reference number Date of hospitalisation Year of birth Gender Main diagnosis: ICD-10 [13] Baseline quality of life: EQ-5D [14–16]
Therapy data	Main intervention: CHOP [17] Date of intervention Complications: Clavien-Dindo Score [12]
Discharge data	Date of discharge Discharge where to (home, rehabilitation, nursing home or other) Patient satisfaction (ANQ questionnaire [18])
Outcome data	Follow-up quality of life: EQ-5D [14–16], comparison with prehospital score (baseline)

ANQ = Swiss National Association for Quality Development in Hospitals and Clinics, Nationaler Verein für Qualitätsentwicklung in Spitälern und Kliniken; AQC= Consortium for Quality Assurance in Surgery; AG für Qualitätssicherung in der Chirurgie; CHOP = Schweizerische Operationsklassifikation; EQ-5D = Euro-Qol-5D; ICD = International Classification of Diseases; SGC-SSC = Swiss Association of Surgeons, Schweizerische Gesellschaft Chirurgie

Published under the copyright license "Attribution – Non-Commercial – No Derivatives 4.0".
No commercial reuse without permission. See http://emh.ch/en/services/permissions.html.

impairment is also a risk factor for fractures [21–24]. In patients with cognitive impairment, the use of instruments based on patient reporting is limited. On the other hand, such instruments often aim at the functional status of patients, and hence it is possible to ask proxies, in particular relatives, friends or nursing home staff [25]. There is research on the agreement between proxy and patient assessments that indicates a worse perception of HRQoL by proxies than by patients [26]. Hence it is unclear to what degree the use of proxies is valid and how the choice between self-reporting and reporting by a proxy should be made.

IV. Individual pre-surgery risk factors

To allow comparison between hospitals and to be able to judge individual outcomes, knowledge about pre-surgery risk factors at the individual level is necessary. Hence a decision is needed as to which risk factors should be included in the MDS and how these risk factors can be assessed in a valid manner. One specific aspect is comorbidities, which are known to play a role in therapeutic success after hip surgery [27–30]. Comorbidities are recorded today as clinical routine; however, this registration is aimed and motivated by the diagnosis related group (DRG)-based payment system. Whether these records are sufficiently valid to allow an individual risk calculation is an open question. Information about comorbidities can also be summarised using various scoring systems, such as the Charlson Comorbidity Index [31, 32] or the Elixhauser Comorbidity Index (ECI) [33], and it has been shown that comorbidity scores based on medication may be a reliable alternative [34–36]

V. Individual peri- and postoperative factors

Peri- and postoperative factors, in particular complications, are typically evaluated as outcomes related to treatment quality. For this simple reason, they should be reflected in the MDS, even if adjustment in hospital comparisons might not be appropriate. Knowledge about complications can also help to understand a limited response in terms of HRQoL in a single patient. However how to report complications in a MDS and how to assess them in clinical routine is an open question.

Rationale and research question

The planned study tries to address the issues mentioned above in the context of hip surgery, covering both acute and elective patients. In these patient groups, the Hip Disability and Osteoarthritis Outcome Score [37–40] (HOOS) is a well-established disease-specific HRQoL assessment instrument. The HOOS is currently introduced in the University Hospital Basel (USB) as part of the ICHOM (International Consortium for Health Outcomes Measurement) initiative [41]. Sit-to-stand and walking-short-distance have been recommended as general activities to be included in a minimum core set of performance-based tests to assess physical function in subjects with hip osteoarthritis [42] and – based on a variety of systematic comparisons [38, 42–45] – the 30s chair-stand test [46–49] and the 40-metre fast-paced walk test [38, 42, 45, 50] have been recommended as specific tests [42]. HUI3 [19] may be criticised in this patient group, as it includes several dimensions (vision, hearing, speech and dexterity) that cannot be expected to improve after hip surgery. It should be noted that all these instruments are validated to assess HRQoL or have been shown to be predictive for HRQoL. However, the question of which of these instruments is best suited to provide outcomes in the context of evaluating treatment quality has yet not been addressed systematically.

With respect to using proxies the study should allow testing of an identification procedure for a proxy in each patient. This way it should be possible to obtain data from a proxy and a self-assessment in patients with sufficient cognitive abilities to perform the latter. A comparison can shed at least some light on the validity of proxy assessments in patients with insufficient cognitive abilities (using the comprehensive Mental Status Questionnaire [MSQ] of Kahn et al [51]). In addition, we can study the usefulness of a clinical assessment of cognitive ability and other factors to develop a rule on when to prefer a proxy assessment to a self-assessment and when to combine the two.

With respect to complications, the study can benefit from a recent initiative to validate and investigate the clinical feasibility of the new CLASSIC system [52] for intraoperative complications, in analogy to the Clavien-Dindo classification [12]. This investigation is performed in a multicentre study, in which the University Hospital Basel is participating. The collection of data on single major complications, as well as the application of several classification systems, is envisioned. With respect to comorbidities, the study should allow the comparison of several alternative assessments, including medication scores, as medication on admission is routinely documented.

In summary, the overall objective of the study is to inform the design of the national Swiss surgical registry with respect to five general research questions, taking into account the specific circumstances just outlined. Consequently, we have the following five specific objectives:

to decide on whether the use of the generic instruments HUI3 and EQ-5D instead of the disease specific instrument HOOS or the 30s chair-stand test and the 40m fast passed walk test can be justified in a register aiming at providing information on treatment quality,

to recommend a specific time-point for a single follow-up assessment in this patient group,

to give some guidance on the use of proxy assessments of HRQoL,

to recommend a set of pre-operative risk factors to be documented,

to make a recommendation on the approach to document complications comparing the CLASSIC, the Clavien-Dindo and the ICH-GCP classification systems as well as documentation of single major complications.

In the present paper we outline the basic design of the study and how we intend to approach the five objectives.

Methods and design

Ethics approval

The study was approved by the Ethikkommission Nordwest- und Zentralschweiz (EKNZ) under reference number 2017-00763.

Published under the copyright license “Attribution – Non-Commercial – No Derivatives 4.0”.
No commercial reuse without permission. See http://emh.ch/en/services/permissions.html.

Design

This study is designed as a longitudinal observational multicentre study. All patients suffering from a femoral neck fracture or from arthritis of the hip joint with indication for partial or total prosthetic joint replacement surgery will be offered participation. The study is based on combining routine data from clinical standard practice with specifically documented data to be reported by the treating clinician or by data collected in cooperation with the patient based on questionnaires and two functional tests. Data will be collected at baseline, during hospitalisation / at discharge and at three routine follow-up visits. All patients will additionally be asked to name a person to serve as a proxy for the assessment of the proxy-perceived HRQoL.

Study population and recruitment

Recruitment started on 15 January 2018. Inclusion and exclusion criteria are summarised in table 2. Refusal of study participation by the designated proxy is not an exclusion criterion for the index patient.

Patients with chronic pathologies and an indication for surgery due to arthritis of the hip joint will be consecutively recruited from the outpatient clinics by the project leaders or their delegates. In cases of screening failures, the reasons for noninclusion will be documented. For the diagnosis of arthritis of the hip joint, a conventional anteroposterior view of the pelvis and an axial view of the affected hip joint are prerequisite. In the presence of relevant symptoms, clinical limitation in the range of movement and/or visible changes in the x-rays following the American College of Rheumatology classification [53], the possibility of prosthetic joint replacement will be discussed with the patient, including a discussion of potential alternatives and possible complications. If the patient consents to hip replacement surgery she/he will be informed about the study and asked to participate.

Patients with acute pathologies (femoral neck fracture) will be consecutively recruited via the emergency departments of the participating centres by the project leaders or their delegates. For the diagnosis of a femoral neck fracture a conventional anteroposterior view of the pelvis and an axial view of the affected hip joint are prerequisite. If these examinations verify the presence of a femoral neck fracture, an indication for surgical treatment has to be posed following current practice. Patients are informed about the therapeutic options, the recommendation of surgical treatment with joint preservation or use of a partial or total hip replacement depending on patient age and demand, as well as about the potential complications of the treatment options. If the patient consents to joint replacement surgery she/he will be informed about the study and asked to participate.

Table 2: Inclusion and exclusion criteria for the MIDAS study.

Inclusion criteria	Presentation at one of the participating centres with one of the following pathologies: • Femoral neck fracture with indication for partial or total hip replacement • Arthritis of the hip joint with indication for prosthetic joint replacement surgery
	Age ≥18 years
	Written informed consent by the patient or a legal representative
Exclusion criteria	Refusal of standard treatment
	Refusal of study participation
	Patients not intending to perform routine follow-up visits at the participating hospitals (e.g., if coming from abroad)
	Known or newly diagnosed malignancy
	Palliative care situation
	Participation during the last 3 months in an interventional clinical trial potentially interacting with the aims of the current study (e.g., trials in musculoskeletal / rheumatologic disease, drug trials influencing the quality of life, etc.)
	Foreign language patients for whom it is unrealistic to obtain the patient-reported outcomes in spite of assistance by a study nurse

A relevant proportion of patients in this study may suffer from dementia or cognitive impairment of various degrees without having a specifically appointed legal guardian or legal representative. These patients will not be excluded from the study. If capability is questionable, or if there is proof of incapability but the patient gives assent to the participation in the study, proxies as described in Article 378 Swiss civil code [54] will be contacted to give informed consent

In both patient groups, all patients willing to participate are asked to name a proxy. Proxies will be equally informed about the study and asked for informed consent. To facilitate the data collection, proxies will be asked for telephone or email contact information to complete the questionnaires in case they do not accompany the patient to visits.

Study variables

All patient-related variables to be collected are summarised in table 3, together with information on the mode and the time–point of collection. Additional information on some of the instruments / classification systems used is provided in table 4. For the ECI, HUI3, EQ-5D and HOOS, all single items will be recorded. Patients will be asked to fill in all questionnaires on their own, but they may ask the study nurse for assistance. The patient-reported outcomes collected at baseline will refer to the pre-fracture status in the patients with an acute fracture. For the housing situation, the following graduation is used:

- independently at home
- at home with occasional professional/familiar support
- at home with regular professional/familiar support (maximum 2/week)
- at home with daily professional/familiar support (once or twice per day)
- at home with constant professional/familiar support (three times per day, in-house nursing)
- institutionalised (level of assisted accommodation)
- institutionalised (level of minor support like meals, logistics, activities)
- institutionalised (dependent)

The family status will be based on the following categories: living alone; living together with a partner; living together with other family members; living together with partner and other family members. Patient satisfaction will be addressed by two questions as suggested by Hamilton et al. [57].

Published under the copyright license “Attribution – Non-Commercial – No Derivatives 4.0”.
No commercial reuse without permission. See http://emh.ch/en/services/permissions.html.

Proxies will be asked for their level of education and their proximity to the patient using the following classification: living together with the proxy; daily contact with the proxy; weekly contact; less than weekly contact. Proxies will be approached by the study nurse at baseline and at one randomly selected follow-up visit to provide data on the HRQoL measures. In the case of difficulties in obtaining information about patient characteristics, the proxies will be involved, too.

The following implant characteristics will be documented: implant type and size (shaft and cup); fixation technique (cemented vs pressfit). However, it should be noted that no comparisons between implant types or implant characteristics are planned, as this is beyond the scope of this study. Such comparisons require a large-scale registry.

Analytical strategy

We will start with an initial data analysis. Distributions of all variables and associations among the outcome measures will be visualised and described by sample statistics and correlation coefficients. Loss to follow up, refusal to fill out a questionnaire or to perform a functional test, and non-response at the item level will be described with respect to the amount and the relation to patient characteristics and time.

All analyses performed will be subjected to several sensitivity analyses. These are outlined in appendix 1.

Several research questions require analysis of the (unadjusted or adjusted) effect of single covariates on scores / change in scores. We will in general perform the following steps to improve comparability across outcomes and covariates: (1) all outcomes are standardised to a population standard deviation of 1; (2) the effect of continuous or ordinal factors is reported with respect to the difference between the 90th percentile and the 10th percentile.

Research question I

We approach the comparison of the instruments by investigating statistical properties of scores derived from these instruments. In order to make a fair comparison, we have to address three basic issues. (1) Which scores do we want to derive from the instruments? (2) How should we assess

Table 3: Patient-related variables to be collected with mode of collection and time-points.

Variable	Mode of data collection	Time-points
Patient characteristics		
Date of hospitalisation	PR	B
Main diagnosis (ICD-10) [13]	PR	B
Age at surgery, gender	PR	B
Education	SN	B
Body mass index	SN	B
Substance utilisation (nicotine, alcohol, drugs)	SN	B
Housing situation and family status	SN	B, F
Use of walking aids	SN	B, F
Distance to hospital	SN	B
Elixhauser Comorbidity Index (ECI) [33]	TC	B
ASA score [55]	TC	B
Cognitive status (Mental Status Questionnaire) [51]	TC	B
Do-not-resuscitate orders	TC	B
Medication on admission (ATC codes) [56]	AU	B
Comorbidities [13] (ICD-10)	AU	B
Surgery-/hospitalisation-related variables		
Implant characteristics	TC	D
Date of intervention, date of discharge, length of hospital stay	PR	D
Major complications: bleeding, infection, dislocation, fractures, and systemic treatment.	PR	D
Complications: CLASSIC [52], ICH scoring [11], Clavien Dindo [12]	TC	D
Discharge destination	TC	D
Adverse events	TC	D
Outcome and follow-up data		
Health Utility Index Mark 3 (HUI3) [19]	SN	B, F
Euro-Qol-5D (EQ-5D) [16]	SN	B, F
Hip Disability and Osteoarthritis Outcome Score (HOOS) [39, 40]	SN	B, F
30 s sit-to-stand test [42]	SN	B*,F†
40 m fast-paced walk test [42]	SN	B*,F†
Exposure to physiotherapy since last visit	SN	F
End of inpatient rehabilitation	SN	F
Patient satisfaction [57]	SN	D,F
Living status/date of death	SN	F
Adverse events [11]	TC	F
Major complications: Infections, dislocation, fractures and systemic treatment	TC	F
Complications: ICH scoring, Clavien-Dindo	TC	F

ASA = American Society of Anesthesiologists; ATC = anatomical-therapeutic-chemical, AU = data extracted automatically from the patient records; B = baseline; D = discharge; F = follow-up visits; ICD = International Classification of Diseases; ICH = International Conference on Harmonisation; PR = data extracted manually from the patient records by the study nurse; SN = data collected by the study nurse in cooperation with the patient; TC = data to be provided by the treating clinician * Only in patients undergoing elective surgery † At one randomly selected follow-up visit

Published under the copyright license "Attribution – Non-Commercial – No Derivatives 4.0".
No commercial reuse without permission. See http://emh.ch/en/services/permissions.html.

the suitability of each score to assess treatment quality? (3) How should we make the comparison?

1. Which scores do we want to derive from the instruments?

   The two generic instruments directly provide summary scores. To focus HUI3 on the dimensions of interest, we will also consider a HUI3-subindex based on the dimensions ambulation, emotion, cognition, and pain. To investigate a potential gain in combining the two generic instruments, we will also consider combined scores based on HUI3/EQ-5D and HUI3-subindex/EQ-5D. The HOOS does not directly provide a summary score, and we will use a principal component analysis to combine the fives subscores. Both functional tests provide a score directly, but we will also consider the combination of both tests, as well as a combination with all HOOS subscores. In addition, the HOOS pain subscore and the HUI3 pain subscore will be compared directly.

2. How should we assess the suitability of each score to assess treatment quality?

   There is no gold standard for the measurement of treatment quality. Hence we cannot just consider the correlation of each score with some treatment quality score. To assess the ability of a score to reflect the quality of treatment, we pursue the following idea: a good measure of treatment quality should be sensitive to (all) known factors influencing treatment success and should reflect these factors in a whole as good as possible. Consequently, R^2-values when fitting a model with a selection of such factors as covariates and the change from baseline in the score as outcome will serve as the main source of information about a score's ability to reflect treatment quality. Factors to be considered as "known" factors with influence on treatment success are all pre-surgery risk factors (cf. RQ IV), all complications (cf. RQ V), the two patient groups, patient satisfaction, length of hospital stay and the baseline score. Complications and patient satisfaction will be handled as time varying covariates allowing only influencing subsequent outcomes.

3. How should we make the comparison?

   To address the question of whether generic instruments can replace disease specific instruments or functional tests, we have to demonstrate that (some of) the scores based on generic instruments are not inferior to the scores based on disease specific instruments or functional measures. Consequently, we have to agree in advance on a noninferiority boundary, i.e. which decrease in R^2 we are willing to accept as being not relevant. We regard a relative decrease by 20% as acceptable.

These comparisons should also result in a recommendation for one score to be used as primary outcome for the other research questions.

In secondary analyses we will perform the same analyses for the time-point specific scores and investigate the association of single factors with the various scores. In addition, we will consider those scales with a well-established minimal clinically important difference (MCID) and compare the agreement in the decision to have a change above or below the MCID [58, 59]. In patients with acute fracture, we will also investigate the agreement with respect to return to the pre-fracture status. This will be approached by

Table 4: Overview of the instruments and classification systems used.

CLASSIC	A general system for the classification of intraoperative complications proposed by Rosenthal et al [52]. It uses the following grading system: Grade 0: No deviation from the ideal intraoperative course Grade 1: Any deviation from the ideal intraoperative course without the need of any additional treatment or intervention Grade 2: Any deviation from the ideal intraoperative course with the need of any additional treatment or intervention not life-threatening and not leading to permanent disability Grade 3: Any deviation from the ideal intraoperative course with the need of any additional treatment or intervention life-threatening and/or leading to permanent disability Grade 4: Any deviation from the ideal intraoperative course with death of the patient
Clavien-Dindo	A general system for the classification of surgical complications suggested by Dindo et al. [12]. It uses the following grading system: Grade I: Any deviation from the normal postoperative course without the need for pharmacological treatment or surgical, endoscopic and radiological interventions. Allowed therapeutic regimens are: drugs as antiemetics, antipyretics, analgesics and diuretics, electrolytes, and physiotherapy. This grade also includes wound infections opened at the bedside. Grade II: Requiring pharmacological treatment with drugs other than those allowed for grade I complications. Blood transfusions and total parenteral nutrition are also included Grade III: Requiring surgical, endoscopic or radiological intervention Grade IIIa: Intervention not under general anaesthesia Grade IIIb: Intervention under general anaesthesia Grade IV: Life-threatening complication (including CNS complications) requiring IC/ICU management Grade Iva: Single organ dysfunction (including dialysis) Grade IVb: Multiorgan dysfunction Grade V: Death of a patient An additional suffix d indicates that the patient suffers from a complication at the time of discharge.
EQ-5D	The EQ-5D [16] has been developed by the EuroQol group. It consists of five items addressing mobility, self-care, usual activities, pain/discomfort, and anxiety/depression, each to be answered on a five-point Lickert scale.
HOOS	The Hip disability and Osteoarthritis Outcome Score (HOOS) [37] consists of five subscales; pain (10 items), other symptoms (5 items), Function in daily living (ADL, 17 items), function in sport and recreation (4 items) and hip related quality of life (4 items). All items are answered on a five-point Likert scale. A score is calculated for each subscale.
ICH scoring	The adverse event scoring according to ICH [11] uses a three level scale: no event – adverse event/reaction – severe adverse event/reaction
HUI3	The Health Utility Index Version 3 (HUI3) [19] addresses eight attributes vision (2 items), hearing (2 items), speech (2 items), ambulation (1 item), dexterity (1 item), emotion (1 item), cognition (2 items), and pain (1 item). Each item has between 4 and 6 response possibilities. For each attribute a classification on a 4, 5, or 6-point scale is computed. These values can be combined to a general health utility.
MSQ	The Mental Status Questionnaire (MSQ) [51] provides a brief description of cognitive functioning. It consists of ten questions and counts the number of correct answers.
ICH = International Conference on Harmonisation	

Published under the copyright license "Attribution – Non-Commercial – No Derivatives 4.0".
No commercial reuse without permission. See http://emh.ch/en/services/permissions.html.

calculation of pairwise kappa coefficients. In order to understand whether different instruments may by sensitive to a different degree to differences in different parts of the underlying scale (e.g., in low performing or high performing patients) we will perform a non-linear, pairwise regression calibration.

Research question II

We regard a fixed time-point to be suitable for a follow up, if the majority of the patients have reached their final level of HRQoL, as this is the level most essential for the patients. Choosing earlier time-points would give an incomplete picture. In addition, we wish to identify the earliest time-point satisfying this condition in order to avoid disturbances by the natural aging of the patient and to maximise response rates. Consequently, we will approach this question by fitting a random effects growth curve model with individual-specific parameters for the growth curve to the time-specific scores of the primary outcome. Based on the estimated distribution of these parameters, we will determine the time-point at which 75% of the patients have reached a stable level. The starting model will be a change point model with a quadratic increase until an individual-specific time-point at which the scores remain constant is reached. If visual inspection of the empirical growth curves suggests another common shape of the curves, the model will be adapted accordingly. Also, the definition of the optimal time-point may be changed accordingly: If many patients already start to deteriorate within the follow-up period, we will aim at the time-point when 75% of the patients have passed the maximum. If many patients do not reach a stable level within the follow-up period, we will aim at the time-point when 75% of the patients have reached the median final change from baseline.

In order to be able to determine such a time-point, it may be a drawback to use the regular follow-up time-points of 3, 6 and 12 months used routinely in the participating hospitals. During the study we will introduce some variation in these time-points by randomising each patient to one of the nine following time-point patterns: 2, 5 and 9 months; 2, 6 and 10 months; 2, 6 and 12 months; 3, 6 and 9 months; 3, 6 and 10 months; 3, 6 and 12 months; 3, 7 and 11 months; 3, 7 and 12 months; 4, 8 and 12 months.

Research question III

In order to make recommendations about the use of proxy assessments, two basic questions have to be addressed. (1) Are proxy assessments sufficiently close to self-assessments? (2) When should we prefer proxy-assessments over self-assessments?

1. Are proxy assessments sufficiently close to self-assessments?

   We have to consider the distribution of the differences in the primary outcome between proxy and self-assessments, which we will describe as histograms and limits of agreement.

2. When should we prefer proxy-assessments over self-assessments?

   We will use regression models to develop a rule to predict raw or absolute differences. If we are able to identify a rule that predicts an absolute difference of more than 0.25 SD of the outcome measure in more than 5% of the patients, we will recommend using this rule in future. The factors to be considered as potential predictors for large differences are the clinical assessment of cognitive ability (MSQ), age, gender, and educational level of the patient and of the proxy, and proximity of the proxy to the patient. Additional analyses will depict agreement at the item level [60].

Research question IV

Preoperative risk factors should be included in the registry if they can be used for case-mix adjustments. Such factors should be predictive for the outcome and potentially varying in distribution across hospitals. Consequently, we will study the effect of single risk factors as well as groups of risk factors with respect to their ability to predict (changes in) the primary outcome. We will also develop a suggestion for a minimal set of relevant risk factors using the Lasso [61]. As single risk factors we will consider age, gender, education, distance to hospital, family status, housing situation, body mass index (BMI), substance abuse, baseline impairment in visual abilities (according to HUI3), MSQ, baseline HRQoL measurements and baseline values of functional tests. Note that many of these risk factors enter existing risk scoring models for morbidity/mortality in hip fracture patients [62]. With respect to comorbidities we will consider several alternatives for incorporation: the single comorbidities collected routinely as well as by the treating clinician, the ECI, the American Society of Anesthesiologists (ASA) score, single comorbidities identified from the medication and a medication score. The latter will be defined after inspection of the distribution of the ATC codes following principles outlined in the literature [63, 64]. The variation of the distribution across the participating hospitals will be described, too.

Research question V

Taking into account that complications have to be part of a minimal data set in any case, we focus on the question of which single complications/complication classifications should be covered. If a complication does not affect the primary outcome, this is an argument to include it. If a complication affects the primary outcome, but this effect is covered already by a classification system, it may be omitted. Hence in a first step we will determine the complications/complication classifications with limited effect on the primary outcome. In a second step, we will determine among those with an effect a parsimonious subset with sufficient prediction of the primary outcome. The statistical criterion for a limited effect is an adjusted R^2 value less than 10%. The criterion for sufficient prediction is a reduction in adjusted R^2 less than 10% compared to a full model with all complications/classification systems.

For each research question, details of the statistical analyses including the handling of missing values, drop outs and death, the choice of transformations, and model checks to be performed will be fixed in corresponding statistical analysis plans.

Sample size considerations

We expect to be able to recruit 300 patients in this study and expect a drop-out rate of 10% at each follow up visit based on experience from previous studies. Power consid-

Published under the copyright license “Attribution – Non-Commercial – No Derivatives 4.0”.
No commercial reuse without permission. See http://emh.ch/en/services/permissions.html.

erations for the different research questions are presented in appendix 2, and suggest sufficient power to address each of the five objectives.

Discussion

To the best of our knowledge, this is the first study explicitly addressing in an empirical manner questions about the design of a national surgical registry. We try to address some questions of central relevance, in particular the use of PROMs to measure HRQoL. The use of PROMs in longitudinal follow-up is still not very common in surgical registries, which typically focus on complication and readmission rates [3, 4]. However, information on long term follow-up in terms of HRQoL is essential for many reasons. Hence we hope that the results of the study on the choice of instruments and timing of a single follow-up measurement will also inspire other registries to take this step forward. In our opinion this can be a useful addition to following general recommendations on how to build up a surgical registry and can, in particular, help to inform the recommended consensus process to agree on a minimal data set [10].

The aim of our study is to support the design of a nationwide cross-disciplinary surgical registry. Consequently, we aimed to design the study in a way mimicking the envisioned data collection process. For example, we decided to ask the patients to fill in all questionnaires on their own with limited assistance, reflecting the situation to be expected in the future. In a pure research context, it would be advisable to offer more assistance. In some points we were forced to deviate from the envisioned process, such as when collecting data at three time-points or when asking all patients to name a proxy. This was necessary to address the research questions of interest.

One crucial point in designing the Swiss surgical registry will be the definition of the population. In designing this study we followed the traditional approach of including all patients who undergo surgery. However, surgical departments are often also responsible for counselling patients about the choice between surgical and nonsurgical treatment, and the adequateness of this counselling is part of their therapy quality. Hence it would be desirable to include in the registry all patients approaching a surgical department with surgical treatment as one option, or to include all patients with a diagnosis associated with surgical treatment options [65].

The proposed study is a multi-purpose study in the sense that it allows five logically independent research questions to be addressed. We expect that the study will actually allow us also to address further research questions not directly related to the design of the Swiss surgical registry. For example, the simultaneous filling in of three questionnaires related to HRQoL, the performance of two functional tests and information on patient satisfaction allows us to investigate dimensionality and interindividual variability of the course of hip surgery patients 1 year after surgery, and to study the conceptual overlap between instruments using canonical correlations. This may even allow us to develop new short instruments covering the information provided by all these instruments. The comprehensive collection of data on pre- and perioperative factors as well as on outcomes allows us to investigate the interrelation. We have also to expect that during our study some of the measures we assess will start to be collected routinely at some centres as part of their internal quality control, and this may be based on new techniques, such as use of tablets. This may give us additional opportunities to study the impact of data collection conditions on the data obtained.

Our study suffers from some limitations. First of all, we only cover elective and acute hip surgery, whereas the Swiss registry should cover all surgical disciplines. Some of our results, such as the relation between proxy- and self-assessment, may be generalisable to other patient groups. However, questions such as the optimal timing and choice of instruments have to be addressed separately for different patient groups. We have also to expect that the two patient groups differ in their typical trajectories. We implicitly assume that these differences are of a quantitative nature and can be fully explained by considering group membership and patient characteristics as covariates. However, we cannot exclude that the answers to our research questions are indeed different for the two patient groups. Furthermore, at some points we were forced to deviate from the envisioned data collection process in order to address the research question of interest. In general, the central question of how to select from several (valid or surrogate-like) instruments the one best assessing treatment quality is somewhat out of the scope of traditional research on HRQoL and the analytical approach chosen may be suboptimal. In particular, we lack an established frame to choose noninferiority margins. Finally, we are still negotiating the participation of several centres, but we have to expect that research-oriented centres will be overrepresented such that we cannot correctly assess the inter-centre variation to be expected in the national registry.

Acknowledgements

We are truly grateful for the support by Prof. Michael Heberer, who initially had the idea of the study; we furthermore thank all members of the clinical research team of the University Hospital Basel: Ilona Ahlborn, Florian Burckhardt, Celine Bürgi and Anna Padiyath for the realisation of data collection tools and critical revisions of the protocol, as well as adjumed.net for the design of a highly functional database.

Disclosure statement

No financial support and no other potential conflict of interest relevant to this article was reported.

References

1 Delaunay C. Registries in orthopaedics. Orthop Traumatol Surg Res. 2015;101(1, Suppl):S69–75. doi: http://dx.doi.org/10.1016/j.otsr.2014.06.029. PubMed.

2 Niederländer C, Wahlster P, Kriza C, Kolominsky-Rabas P. Registries of implantable medical devices in Europe. Health Policy. 2013;113(1-2):20–37. doi: http://dx.doi.org/10.1016/j.healthpol.2013.08.008. PubMed.

3 Pugely AJ, Martin CT, Harwood J, Ong KL, Bozic KJ, Callaghan JJ. Database and Registry Research in Orthopaedic Surgery: Part 2: Clinical Registry Data. J Bone Joint Surg Am. 2015;97(21):1799–808. doi: http://dx.doi.org/10.2106/JBJS.O.00134. PubMed.

4 Alluri RK, Leland H, Heckmann N. Surgical research using national databases. Ann Transl Med. 2016;4(20):393. doi: http://dx.doi.org/10.21037/atm.2016.10.49. PubMed.

5 Dy CJ, Bumpass DB, Makhni EC, Bozic KJ; AAOS Washington Health Policy Fellowship. The Evolving Role of Clinical Registries: Existing Practices and Opportunities for Orthopaedic Surgeons. J Bone Joint Surg Am. 2016;98(2):. doi: http://dx.doi.org/10.2106/JBJS.O.00494. PubMed.

6 Sebastian AS. Database Research in Spine Surgery. Clin Spine Surg. 2016;29(10):427–9. doi: http://dx.doi.org/10.1097/BSD.0000000000000464. PubMed.

Published under the copyright license “Attribution – Non-Commercial – No Derivatives 4.0”.
No commercial reuse without permission. See http://emh.ch/en/services/permissions.html.

7 Inacio MCS, Paxton EW, Dillon MT. Understanding Orthopaedic Registry Studies: A Comparison with Clinical Studies. J Bone Joint Surg Am. 2016;98(1):. doi: http://dx.doi.org/10.2106/JBJS.N.01332. PubMed.

8 Stey AM, Russell MM, Ko CY, Sacks GD, Dawes AJ, Gibbons MM. Clinical registries and quality measurement in surgery: a systematic review. Surgery. 2015;157(2):381–95. doi: http://dx.doi.org/10.1016/j.surg.2014.08.097. PubMed.

9 Hoffman RL, Bartlett EK, Medbery RL, Sakran JV, Morris JB, Kelz RR. Outcomes registries: an untapped resource for use in surgical education. J Surg Educ. 2015;72(2):264–70. doi: http://dx.doi.org/10.1016/j.jsurg.2014.08.014. PubMed.

10 Mandavia R, Knight A, Phillips J, Mossialos E, Littlejohns P, Schilder A. What are the essential features of a successful surgical registry? a systematic review. BMJ Open. 2017;7(9):. doi: http://dx.doi.org/10.1136/bmjopen-2017-017373. PubMed.

11 International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline. Guideline for good clinical practice E6(R1).

12 Dindo D, Demartines N, Clavien P-A. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. Ann Surg. 2004;240(2):205–13. doi: http://dx.doi.org/10.1097/01.sla.0000133083.54934.ae. PubMed.

13 World Health Organisation. ICD-10 Version: 2016. http://apps.who.int/classifications/icd10/browse/2016. Accessed Nov 14, 2017

14 Brooks RG. 28 Years of the EuroQol Group : An Overview. EuroQol Working Paper Series 15003. 2015. Available from: https://euroqol.org/wp-content/uploads/working_paper_series/EuroQol_Working_Paper_Series_Manuscript_15003_-_Richard_Brooks.pdf.

15 Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20(10):1727–36. doi: http://dx.doi.org/10.1007/s11136-011-9903-x. PubMed.

16 EQ-5D-5L user guide. Basic Information on how to use EQ-5D-5L Instrumemnts. *2015*. https://euroqol.org/wp-content/uploads/2016/09/EQ-5D-5L_UserGuide_2015.pdf. Accessed November 14, 2017.

17 Eidgenössisches Department des Inneren - Bundesamt für Statistik. Schweizerische Operationsklassifikation (CHOP). Systematisches Verzeichnis 2017.

18 Patientenzufriedenheit ANQ. http://www.anq.ch/akutsomatik/patienten-befragung/. Accessed August 27, 2017.

19 Horsman J, Furlong W, Feeny D, Torrance G. The Health Utilities Index (HUI): concepts, measurement properties and applications. Health Qual Life Outcomes. 2003;1(1):54. doi: http://dx.doi.org/10.1186/1477-7525-1-54. PubMed.

20 Van Beeck EF, Larsen CF, Lyons RA, Meerding WJ, Mulder S, Essink-Bot ML. Guidelines for the conduction of follow-up studies measuring injury-related disability. J Trauma. 2007;62(2):534–50. doi: http://dx.doi.org/10.1097/TA.0b013e31802e70c7. PubMed.

21 Shaw FE. Falls in cognitive impairment and dementia. Clin Geriatr Med. 2002;18(2):159–73. doi: http://dx.doi.org/10.1016/S0749-0690(02)00003-4. PubMed.

22 Muir SW, Gopaul K, Montero Odasso MM. The role of cognitive impairment in fall risk among older adults: a systematic review and meta-analysis. Age Ageing. 2012;41(3):299–308. doi: http://dx.doi.org/10.1093/ageing/afs012. PubMed.

23 Liu-Ambrose TY, Ashe MC, Graf P, Beattie BL, Khan KM. Increased risk of falling in older community-dwelling women with mild cognitive impairment. Phys Ther. 2008;88(12):1482–91. doi: http://dx.doi.org/10.2522/ptj.20080117. PubMed.

24 Segev-Jacubovski O, Herman T, Yogev-Seligmann G, Mirelman A, Giladi N, Hausdorff JM. The interplay between gait, falls and cognition: can cognitive therapy reduce fall risk? Expert Rev Neurother. 2011;11(7):1057–75. doi: http://dx.doi.org/10.1586/ern.11.69. PubMed.

25 Pickard AS, Knight SJ. Proxy evaluation of health-related quality of life: a conceptual framework for understanding multiple proxy perspectives. Med Care. 2005;43(5):493–9. doi: http://dx.doi.org/10.1097/01.mlr.0000160419.27642.a8. PubMed.

26 Magaziner J, Simonsick EM, Kashner TM, Hebel JR. Patient-proxy response comparability on measures of patient health and functional status. J Clin Epidemiol. 1988;41(11):1065–74. doi: http://dx.doi.org/10.1016/0895-4356(88)90076-5. PubMed.

27 Mathew RO, Hsu W-H, Young Y. Effect of comorbidity on functional recovery after hip fracture in the elderly. Am J Phys Med Rehabil. 2013;92(8):686–96. doi: http://dx.doi.org/10.1097/PHM.0b013e318282bc67. PubMed.

28 Kabboord AD, van Eijk M, Fiocco M, van Balen R, Achterberg WP. Assessment of Comorbidity Burden and its Association With Functional Rehabilitation Outcome After Stroke or Hip Fracture: A Systematic Review and Meta-Analysis. J Am Med Dir Assoc. 2016;17(11):1066.e13–21. doi: http://dx.doi.org/10.1016/j.jamda.2016.07.028. PubMed.

29 Peter WF, Dekker J, Tilbury C, Tordoir RL, Verdegaal SH, Onstenk R, et al. The association between comorbidities and pain, physical function and quality of life following hip and knee arthroplasty. Rheumatol Int. 2015;35(7):1233–41. doi: http://dx.doi.org/10.1007/s00296-015-3211-7. PubMed.

30 Günther KP, Haase E, Lange T, Kopkow C, Schmitt J, Jeszenszky C, et al. Persönlichkeitsprofil und Komorbidität: Gibt es den "schwierigen Patienten" in der primären Hüftendoprothetik? [Personality and comorbidity: are there "difficult patients" in hip arthroplasty?]. Orthopade. 2015;44(7):555–65 . http://dx.doi.org/10.1007/s00132-015-3097-9. PubMed.

31 Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373–83. doi: http://dx.doi.org/10.1016/0021-9681(87)90171-8. PubMed.

32 Charlson M, Szatrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. J Clin Epidemiol. 1994;47(11):1245–51. doi: http://dx.doi.org/10.1016/0895-4356(94)90129-5. PubMed.

33 Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care. 1998;36(1):8–27. doi: http://dx.doi.org/10.1097/00005650-199801000-00004. PubMed.

34 Clark DO, Von Korff M, Saunders K, Baluch WM, Simon GE. A chronic disease score with empirically derived weights. Med Care. 1995;33(8):783–95. doi: http://dx.doi.org/10.1097/00005650-199508000-00004. PubMed.

35 Perkins AJ, Kroenke K, Unützer J, Katon W, Williams JW, Jr, Hope C, et al. Common comorbidity scales were similar in their ability to predict health care costs and mortality. J Clin Epidemiol. 2004;57(10):1040–8. doi: http://dx.doi.org/10.1016/j.jclinepi.2004.03.002. PubMed.

36 Schneeweiss S, Maclure M. Use of comorbidity scores for control of confounding in studies using administrative databases. Int J Epidemiol. 2000;29(5):891–8. doi: http://dx.doi.org/10.1093/ije/29.5.891. PubMed.

37 Nilsdotter A, Bremander A. Measures of hip function and symptoms: Harris Hip Score (HHS), Hip Disability and Osteoarthritis Outcome Score (HOOS), Oxford Hip Score (OHS), Lequesne Index of Severity for Osteoarthritis of the Hip (LISOH), and American Academy of Orthopedic Surgeons (AAOS) Hip and Knee Questionnaire. Arthritis Care Res (Hoboken). 2011;63(S11, Suppl 11):S200–7. doi: http://dx.doi.org/10.1002/acr.20549. PubMed.

38 Kennedy DM, Stratford PW, Wessel J, Gollish JD, Penney D. Assessing stability and change of four performance measures: a longitudinal study evaluating outcome following total hip and knee arthroplasty. BMC Musculoskelet Disord. 2005;6(1):3. doi: http://dx.doi.org/10.1186/1471-2474-6-3. PubMed.

39 Roos EM. HOOS. http://www.koos.nu. Accessed Nov 14, 2017.

40 Blasimann A, Dauphinee SW, Staal JB. Translation, cross-cultural adaptation, and psychometric properties of the German version of the hip disability and osteoarthritis outcome score. J Orthop Sports Phys Ther. 2014;44(12):989–97. doi: http://dx.doi.org/10.2519/jospt.2014.4994. PubMed.

41 ICHOM. http://www.ichom.org. Accessed August 28, 2017.

42 Dobson F, Hinman RS, Roos EM, Abbott JH, Stratford P, Davis AM, et al. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. Osteoarthritis Cartilage. 2013;21(8):1042–52. doi: http://dx.doi.org/10.1016/j.joca.2013.05.002. PubMed.

43 Dobson F, Hinman RS, Hall M, Terwee CB, Roos EM, Bennell KL. Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: a systematic review. Osteoarthritis Cartilage. 2012;20(12):1548–62. doi: http://dx.doi.org/10.1016/j.joca.2012.08.015. PubMed.

44 Bennell K, Dobson F, Hinman R. Measures of physical performance assessments: Self-Paced Walk Test (SPWT), Stair Climb Test (SCT), Six-Minute Walk Test (6MWT), Chair Stand Test (CST), Timed Up & Go (TUG), Sock Test, Lift and Carry Test (LCT), and Car Task. Arthritis Care Res (Hoboken). 2011;63(S11, Suppl 11):S350–70. doi: http://dx.doi.org/10.1002/acr.20538. PubMed.

45 Steffen TM, Hacker TA, Mollinger L. Age- and gender-related test performance in community-dwelling elderly people: Six-Minute Walk Test, Berg Balance Scale, Timed Up & Go Test, and gait speeds. Phys Ther. 2002;82(2):128–37. doi: http://dx.doi.org/10.1093/ptj/82.2.128. PubMed.

46 Jones CJ, Rikli RE, Beam WCA. A 30-s chair-stand test as a measure of lower body strength in community-residing older adults. Res Q Exerc Sport. 1999;70(2):113–9. doi: http://dx.doi.org/10.1080/02701367.1999.10608028. PubMed.

Published under the copyright license "Attribution – Non-Commercial – No Derivatives 4.0".
No commercial reuse without permission. See http://emh.ch/en/services/permissions.html.

47 Millor N, Lecumberri P, Gómez M, Martínez-Ramírez A, Izquierdo M. An evaluation of the 30-s chair stand test in older adults: frailty detection based on kinematic parameters from a single inertial unit. J Neuroeng Rehabil. 2013;10(1):86. doi: http://dx.doi.org/10.1186/1743-0003-10-86. PubMed.

48 Gill S, McBurney H. Reliability of performance-based measures in people awaiting joint replacement surgery of the hip or knee. Physiother Res Int. 2008;13:141–52.

49 Unver B, Kahraman T, Kalkan S, Yuksel E, Karatosun V, Gunal I. Test-retest reliability of the 50-foot timed walk and 30-second chair stand test in patients with total hip arthroplasty. Acta Orthop Belg. 2015;81(3):435–41. PubMed.

50 Kennedy D, Stratford PW, Pagura SM, Walsh M, Woodhouse LJ. Comparison of gender and group differences in self-report and physical performance measures in total hip and knee arthroplasty candidates. J Arthroplasty. 2002;17(1):70–7. doi: http://dx.doi.org/10.1054/arth.2002.29324. PubMed.

51 Kahn RL, Goldfarb AI, Pollack M, Peck A. Brief objective measures for the determination of mental status in the aged. Am J Psychiatry. 1960;117(4):326–8. doi: http://dx.doi.org/10.1176/ajp.117.4.326. PubMed.

52 Rosenthal R, Hoffmann H, Clavien PA, Bucher HC, Dell-Kuster S. Definition and classification of intraoperative complications (CLASSIC): Delphi study and pilot evaluation. World J Surg. 2015;39(7):1663–71. doi: http://dx.doi.org/10.1007/s00268-015-3003-y. PubMed.

53 Altman R, Alarcón G, Appelrouth D, Bloch D, Borenstein D, Brandt K, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis Rheum. 1991;34(5):505–14. doi: http://dx.doi.org/10.1002/art.1780340502. PubMed.

54 Bundesversammlung der Schweizer Eidgenossenschaft: Schweizerisches Zivilgesetzbuch. (Stand 1.9.2017).

55 Saklad M. Grading for Patients for Surgical Procedures. Anesthesiology. 1941;2(3):281–4. doi: http://dx.doi.org/10.1097/00000542-194105000-00004.

56 Codes ATC. http://www.atccode.com/. Accessed November 14, 2017.

57 Hamilton DF, Lane JV, Gaston P, Patton JT, Macdonald DJ, Simpson AH, et al. Assessing treatment outcomes using a single question: the net promoter score. Bone Joint J. 2014;96-B(5):622–8. doi: http://dx.doi.org/10.1302/0301-620X.96B5.32434. PubMed.

58 Luo N, Johnson J, Coons SJ. Using instrument-defined health state transitions to estimate minimally important differences for four preference-based health-related quality of life instruments. Med Care. 2010;48(4):365–71. doi: http://dx.doi.org/10.1097/MLR.0b013e3181c162a2. PubMed.

59 Paulsen A, Roos EM, Pedersen AB, Overgaard S. Minimal clinically important improvement (MCII) and patient-acceptable symptom state (PASS) in total hip arthroplasty (THA) patients 1 year postoperatively. Acta Orthop. 2014;85(1):39–48. doi: http://dx.doi.org/10.3109/17453674.2013.867782. PubMed.

60 Farin E. Integration of patient and provider assessments of mobility and self-care resulted in unidimensional item-response theory scales. J Clin Epidemiol. 2009;62(10):1075–84. doi: http://dx.doi.org/10.1016/j.jclinepi.2008.11.014. PubMed.

61 Tibshirani R. Regression Selection and Shrinkage via the Lasso. J R Stat Soc B. 1996;58(1):267–88. doi: http://dx.doi.org/10.2307/2346178.

62 Marufu TC, Mannings A, Moppett IK. Risk scoring models for predicting peri-operative morbidity and mortality in people with fragility hip fractures: Qualitative systematic review. Injury. 2015;46(12):2325–34. doi: http://dx.doi.org/10.1016/j.injury.2015.10.025. PubMed.

63 Huber CA, Szucs TD, Rapold R, Reich O. Identifying patients with chronic conditions using pharmacy data in Switzerland: an updated mapping approach to the classification of medications. BMC Public Health. 2013;13(1):1030. doi: http://dx.doi.org/10.1186/1471-2458-13-1030. PubMed.

64 Cortaredona S, Pambrun E, Verdoux H, Verger P. Comparison of pharmacy-based and diagnosis-based comorbidity measures from medical administrative data. Pharmacoepidemiol Drug Saf. 2017;26(4):402–11. doi: http://dx.doi.org/10.1002/pds.4146. PubMed.

65 Cook JA, McCulloch P, Blazeby JM, Beard DJ, Marinac-Dabic D, Sedrakyan A; IDEAL Group. IDEAL framework for surgical innovation 3: randomised controlled trials in the assessment stage and evaluations in the long term study stage. BMJ. 2013;346(jun18 3):f2820. doi: http://dx.doi.org/10.1136/bmj.f2820. PubMed.

66 Van Balen R, Essink-Bot ML, Steyerberg E, Cools H, Habbema DF. Quality of life after hip fracture: a comparison of four health status measures in 208 patients. Disabil Rehabil. 2003;25(10):507–19. doi: http://dx.doi.org/10.1080/0963828031000090443. PubMed.

Published under the copyright license “Attribution – Non-Commercial – No Derivatives 4.0”.
No commercial reuse without permission. See http://emh.ch/en/services/permissions.html.

Appendix 1

Sensitivity analyses

Patient satisfaction, participation in rehabilitation and exposure to physiotherapy are intermediate variables on the pathway from surgery to medium-term HRQoL. Consequently, risk factor analyses should not be adjusted for such factors. Investigating their role as mediators is not an aim of this project. However, for assessing the relevance of the results of this project with respect to the future use of the envisioned, prospective data collection, it would be good to know that the results do not substantially differ in subgroups defined by these variables. Hence we will conduct sensitivity analyses restricting the population to patients with or without rehabilitation and consider subgroups differing in exposure to physiotherapy or patient satisfaction.

The EQ5D-5L includes also a visual analogue scale. We will compare the markings of the patients with their results on the questionnaire part of the EQ5D-5L. Distinct discrepancies may give us a hint about the questionnaire data, which may be invalid. In a sensitivity analysis, we will investigate whether these observations have an undesirable, strong influence on our results.

In a further sensitivity analysis we will investigate the stability of the results over the two patient groups, whenever a joint analysis has been reported.

Appendix 2

Power considerations for the different research questions

Our sample size considerations are based on the results of a simulation study. In this study we consider a specific data generating model for the outcome scores. In the model we assume that we have 10 independent, binary risk factors each with a prevalence of 1/3 and standardised to mean 0. The average effect of the covariates is denoted by β and the effect of the single covariates is equidistantly spread between $2/11 \times \beta$ and $20/11 \times \beta$, i.e. from a very small effect close to 0 up to an effect nearly twice as β. The effects are assumed to be constant over time, but reduced to 30% for the baseline measurement.

For a single score y, the model reads for patient i at time-point t:

$$y(i,t) = \mu(i,t) + \beta_1 x_1(i) + ... + \beta_{10} x_{10}(i) + \varepsilon(i,t)$$

Here μ(i,t) represents the growth pattern of patient i. We assume that all patients follow a quadratic model starting at baseline with a patient specific level α(i) and reaching the maximum value γ(i) at time p(i). We assume α(i) to be uniformly distributed between 20 and 60, γ(i) to be uniformly distributed between 60 and 100, and p(i) uniformly distributed between 9 and 13 months. All three parameters are drawn independently from each other, such that we have a wide variation of individual growth patterns from a nearly constant curve up to curves with a steep increase. The noise in the data which we cannot explain by the individual growth pattern and the covariates is described by ε(i,t), which is assumed to be independently normally distributed with a standard deviation of 15. We assume in our simulations that we have complete data at baseline from 300 patients and allow a 10% drop out rate at each time-point, such that about 72% of all patients provide data for the final time-point.

Research question (RQ) II requires analysis of the raw scores. The other research questions require analysis of the change scores, at least in the primary analysis. These follow again a linear model, but the regression coefficients are reduced by 30%. We refer to these values by β*. In the sequel our considerations are based on analysis of the change scores by fitting a regression model for each time-point and then averaging the regression coefficients over the three time-points (RQIV, V), or averaging the log-transformed ratios of the adjusted R^2 values between two scores. Inference in the latter case is based on a non-parametric bootstrap at the patient level. The 95% CIs (confidence intervals) are then back transformed to the ratio scale.

RQI: The key situation for sample size considerations is the case of two scores with the same true R^2 value. The ratio of the true R^2 values is then 1.0, and we have to demonstrate that the standard error of the estimated ratio is small enough to ensure that the lower bound of a 95% confidence interval for the ratio is above 0.8. The probability to reach this aim depends mainly on two factors: First on the magnitude of the R^2 values, and second on the degree of correlation between the scores. With respect to the first issue we refer to a paper by van Balen et al [66], who studied the

Published under the copyright license "Attribution – Non-Commercial – No Derivatives 4.0".
No commercial reuse without permission. See http://emh.ch/en/services/permissions.html.

explained variation in HRQoL four months after hip fracture by a model with the four variables "living in a home for the elderly", "number of comorbidities", "age at hospital admission" and "MMSE-score 1 week after hospital admission". Using the Nottingham Health profile (NHP) as outcome, they obtained adjusted R^2 values of 0.37. Using the Rehabilitation Activities Profile (RAP), they obtained an adjusted R^2 value of 0.58. Although our situation is not completely comparable (change scores instead of follow-up values, measurements at several time-points, more than 4 risk factors available, different instruments), we believe that these results indicate that we can also expect R^2 values in this range.

With respect to the second issue, we vary three elements of the correlation between the two scores in our simulation: the correlation between the growth curves, measured by the Spearman correlation of the two score specific α values and the two score specific γ values in each patient (ρ_1), the correlation between the error terms (ρ_2) and whether the covariates have the same effect, or different effects. For the latter we consider the special case that the effects are reversed in their order. For the first score the first covariate had the smallest effect and the last covariate the largest effect, and for the second score it is just the other way round.

We consider two different choices for the average effect β, namely 19 and 24, resulting in R^2 values within the span mentioned above. From the simulation we report here the mean observed R^2 values (R^2), the standard deviation of the estimated ratios (se, as these values correspond to the standard error of the ratio estimate), and the frequency to have a lower bound above 0.8 of the 95% CI for the ratio (power).

The following table presents the results for ρ_1=0.3, i.e. only a moderate correlation of the growth patterns. The results for ρ_1=0.5 were only slightly better.

The results suggest that we have a reasonable power to reach our aim (i.e. to demonstrate a ratio above 0.8), if we have succeeded in selecting the potential factors in a way to reach R^2 values above 0.45 and if the two different scores depend to a similar degree on the single factors when adjusted for all other factors.

RQII: Here we base the sample size considerations on an attempt to estimate the upper 75% percentile of the distribution of the individual peaks in the quadratic model by fitting a random effects quadratic model to the individual growth curves under the assumption of a joint normal distribution for the three parameters of the quadratic model. The 75% quantile can then be determined based on the parameter estimates. Since we are considering now only one score, the precision of the estimate depends only on the value of β, i.e. the true average effect. For both β=19 and β=24 we observe standard errors in the magnitude of 1.2 for the estimated 75% quantile. This suggests that we can estimate the position of the peak with a precision which allow us to determine an adequate time-point for the future follow-up.

RQIV/V: These research questions focus mainly on estimating the effect of single factors in multiple regression models. We could observe standard errors in the magnitude of 2.8 for all regression coefficients for both choices of β. Since the average true effects in our simulation to be 13.3 and 16.8, respectively, this suggests a reasonable power to distinguish between factors with small effects and factors with large effects.

RQIII: With respect to analysis of the difference between self-assessment and proxy-assessment, we have to take into account that we can only use those patients for whom we can obtain both values and that we have at most one pair of such values for the change score in each patient. We expect that this will be the case in 50% of the patients. With 150 patients, we can estimate the standard deviation of these differences with a standard error corresponding to 6% of the standard deviation, suggesting that we have a sufficient precision to describe the variation of these differences.

				Reversed effects		Equal effects	
β	**β***	**ρ_2**	**R^2**	**SE**	**Power (%)**	**SE**	**Power (%)**
19	13.3	0.3	0.41	0.116	44	0.094	59
		0.6		0.105	50	0.081	72
24	16.8	0.3	0.53	0.081	72	0.067	88
		0.6		0.078	78	0.057	94

Published under the copyright license "Attribution – Non-Commercial – No Derivatives 4.0".
No commercial reuse without permission. See http://emh.ch/en/services/permissions.html.