Individual perspectives on outcomes in Diabetes

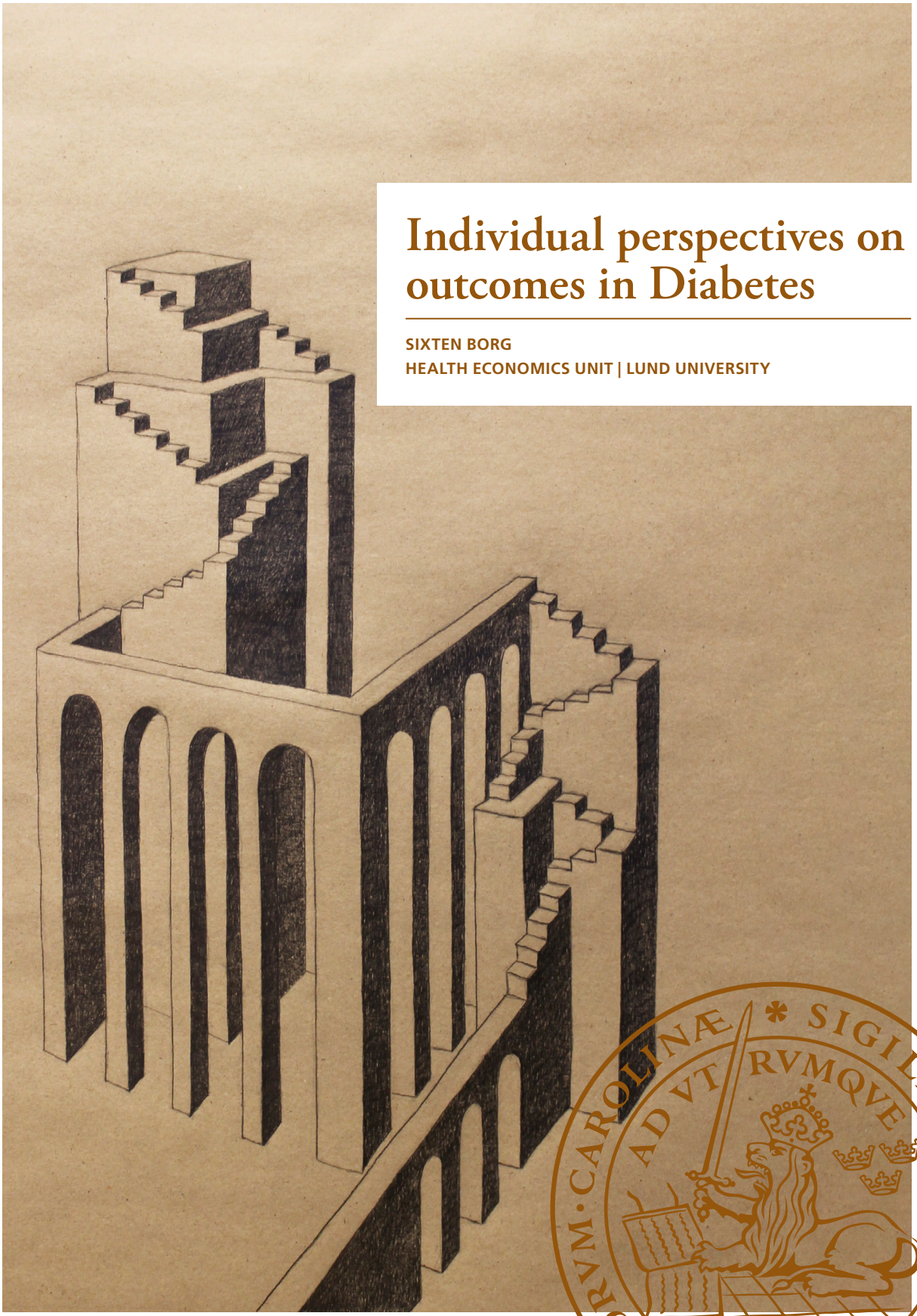
Borg, Sixten

2020

Document Version:
Publisher's PDF, also known as Version of record

Individual perspectives on outcomes in Diabetes

SIXTEN BORG
HEALTH ECONOMICS UNIT | LUND UNIVERSITY



Printed by Media-Tryck, Lund 2020 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

FACULTY OF MEDICINE

Department of Clinical Sciences in Malmö
Health Economics Unit

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2020:52
ISBN 978-91-7619-913-8
ISSN 1652-8220

Individual perspectives on outcomes in Diabetes

Individual perspectives on outcomes in Diabetes

Sixten Borg

LUND
UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty of Medicine, Lund University, Sweden.
To be defended at Seminar Room "37:an", CRC, Malmö on 14 May 2020 at 13:15.

Faculty opponent
Professor Unto Häkkinen

Organization LUND UNIVERSITY	**Document name** Doctoral dissertation
	Date of issue
Author(s) Sixten Borg	Sponsoring organization

Title and subtitle Individual perspectives on outcomes in Diabetes

Abstract

Purpose:
The purpose was to promote an individual perspective in diabetes care through developing how patient-reported measurements are used in the evaluation of the situation of an individual with diabetes, how diabetes care works together with the individual, and in the improvement of diabetes care.

Methods and results:
In study I, we identified procedures for developing measurement scales for a questionnaire, examine their measurement properties and compute scale scores. In study II, we used these procedures on an updated version of the questionnaire. We obtained scales for quantifying wellbeing, abilities, freedom from worries and barriers, and judgments of experience of diabetes care were developed. They had acceptable measurement properties and could be used to describe our study populations, compare groups, and identify vulnerable individuals. In study III we examined a method for estimating the quality of life in type 1 diabetes, based on our scales together with risk factors for diabetes complications, e.g. HbA1c. We could measure the quality of life, and the measure allowed every individual to use their own importance weights for the variables involved. The method also estimated the individual's improvement potential. In study IV, we tried to identify predictors of future costs and future risk factors, among our scales. Ability to manage diabetes predicted HbA1c in type 1 diabetes. Satisfaction with treatment predicted HbA1c in type 2 diabetes.

Conclusions:
We have contributed to the individual perspective in diabetes in several ways. We could quantify patient-reported outcomes and experience measures. They could be used with risk factors for diabetes complications, to describe an individual's situation, to estimate the quality of life, and for predicting future HbA1c. Taken together, this could be used for developing and improving diabetes care and the situation for an individual with diabetes, with regards to clinical practice, relevant outcomes and their valuation.

Key words Diabetes. Patient-Reported Outcomes Measures. Patient-Reported Experience Measures. Individual perspective.

Classification system and/or index terms (if any)

Supplementary bibliographical information	**Language** English
ISSN and key title 1652-8220 Individual perspectives on outcomes in Diabetes	**ISBN** 978-91-7619-913-8 978-91-7619-913-8

Recipient's notes	**Number of pages** 62+appendices	Price
	Security classification	

Individual perspectives on outcomes in Diabetes

Sixten Borg

LUND
UNIVERSITY

Cover art by Sixten Borg, *Nio års hattande*

Health Economics Unit
Department of Clinical Sciences in Malmö
Lund University

ISSN 1652-8220
ISBN 978-91-7619-913-8
Lund University, Faculty of Medicine Doctoral Dissertation Series 2020:52

Printed in Sweden by Media-Tryck, Lund University
Lund 2020

"Har du fått på öronen idag?"

Table of Contents

Abstract

Purpose:
The purpose was to promote an individual perspective in diabetes care through developing how patient-reported measurements are used in the evaluation of the situation of an individual with diabetes, how diabetes care works together with the individual, and in the improvement of diabetes care.

Methods and results:
In study I, we identified procedures for developing measurement scales for a questionnaire, examine their measurement properties and compute scale scores. In study II, we used these procedures on an updated version of the questionnaire. We obtained scales for quantifying wellbeing, abilities, freedom from worries and barriers, and judgments of experience of diabetes care were developed. They had acceptable measurement properties and could be used to describe our study populations, compare groups, and identify vulnerable individuals. In study III we examined a method for estimating the quality of life in type 1 diabetes, based on our scales together with risk factors for diabetes complications, e.g. HbA1c. We could measure the quality of life, and the measure allowed every individual to use their own importance weights for the variables involved. The method also estimated the individual's improvement potential. In study IV, we tried to identify predictors of future costs and future risk factors, among our scales. Ability to manage diabetes predicted HbA1c in type 1 diabetes. Satisfaction with treatment predicted HbA1c in type 2 diabetes.

Conclusions:
We have contributed to the individual perspective in diabetes in several ways. We could quantify patient-reported outcomes and experience measures. They could be used with risk factors for diabetes complications, to describe an individual's situation, to estimate the quality of life, and for predicting future HbA1c. Taken together, this could be used for developing and improving diabetes care and the situation for an individual with diabetes, with regards to clinical practice, relevant outcomes and their valuation.

Sammanfattning på svenska

Syfte:
Syftet med avhandlingen var att främja det individuella perspektivet i diabetesvården genom att utveckla patientrapporterade mått, och hur de användas för att bedöma en persons situation, för hur vården kan arbeta med enkäten tillsammans med personen, och för hur diabetesvården kan utvecklas.

Metoder och resultat:
Studie I etablerade en ansats för att utveckla en enkäts mätskalor, bedöma deras mätegenskaper, och göra kvantitativa mätningar. Studie II tillämpade ansatsen på en mera utvecklad version av enkäten. Vi lyckades utveckla fungerande skalor för patientrapporterade mått och med dem kvantitativt mäta välmående, förmågor, frihet från oro och hinder, och bedömningar av upplevelsen av diabetesvården. Skalorna kunde användas för att beskriva våra studiepopulationer, jämföra grupper och identifiera personer med till exempel dåliga utfall. I studie III prövade vi om en metod kunde användas för att mäta livskvalitet bland personer med typ 1-diabetes, med hjälp av våra patientrapporterade mått och riskfaktorer för diabeteskomplikationer, t.ex. HbA1c. Vi fann att metoden fungerade och lyckades mäta livskvaliteten på ett sätt som gjorde att personerna själva kunde styra hur viktiga olika variabler var för deras livskvalitet. Ansatsen kunde också användas för att visa orsaker till livskvalitetsförluster och skatta en persons förbättringspotential. I studie IV undersökte vi om våra patientrapporterade mått predikterade framtida kostnader för diabetesvård och framtida riskfaktorer för diabeteskomplikationer. Förmågan att hantera diabetes predikterade framtida HbA1c i typ 1-diabetes, och nöjdhet med behandling och tekniska hjälpmedel predikterade HbA1c i typ 2-diabetes.

Slutsatser:
Vi har bidragit till det individuella perspektivet i vården på flera sätt. Vi kunde kvantitativt mäta patientrapporterade utfallsmått och patientrapporterade erfarenhetsmått. Dessa kunde tillsammans med riskfaktorer för diabeteskomplikationer användas för att beskriva en persons situation, mäta livskvalitet och prediktera framtida HbA1c. Tillsammans utgör detta underlag för utformning av relevanta åtgärder för att utveckla och förbättra diabetesvården och situationen för en person med diabetes, med hänsyn till samarbetet mellan vården och personen med diabetes, relevanta utfall och deras värdering.

List of papers

Paper I. Borg, S., B. Palaszewski, U. G. Gerdtham, O. Fredrik, P. Roos and S. Gudbjornsdottir (2014). "Patient-reported outcome measures and risk factors in a quality registry: a basis for more patient-centered diabetes care in Sweden." International journal of environmental research and public health 11(12): 12223-12246.

Paper II. Borg, S., K. Eeg-Olofsson, B. Palaszewski, M. Svedbo Engstrom, U. G. Gerdtham and S. Gudbjornsdottir (2019). "Patient-reported outcome and experience measures for diabetes: development of scale models, differences between patient groups and relationships with cardiovascular and diabetes complication risk factors, in a combined registry and survey study in Sweden." BMJ Open 9(1): e025033.

Paper III: Sixten Borg, Ulf-G Gerdtham, Katarina Eeg-Olofsson, Bo Palaszewski, Soffia Gudbjörnsdottir. Quality of life in chronic conditions using patient-reported measures and biomarkers: a DEA analysis in type 1 Diabetes. Health Econ Rev. 2019 Nov 7;9(1):31.

Manuscript IV. Sixten Borg, Ulf-G. Gerdtham, Bo Palaszewski, Katarina Eeg-Olofsson, Ann-Marie Svensson, Soffia Gudbjörnsdottir. Patient-reported measures and the relationships with costs and risk factors for complications in diabetes. In preparation.

Abbreviations and terms

DEA	Data Envelopment Analysis
HbA1c	Glycated haemoglobin
IRT	Item Response Theory
Item	Question (the term item is preferred in the field of IRT)
LDL	Low-density Lipoprotein
NDR	Swedish National Diabetes Register (Nationella Diabetesregistret)
PREM	Patient-Reported Experience Measures
PROM	Patient-Reported Outcomes Measures
QALY	Quality-Adjusted Life Year
SBP	Systolic Blood Pressure
Scale	A group of questions addressing the same underlying concept

1. Introduction

Background

Diabetes is a chronic disease that causes great burden. In the adult population in Sweden, about 5.5% have diabetes [1]. The disease is subdivided into two main types. In type 1, the individual cannot produce insulin and insulin must be taken daily by multiple injections or with an insulin pump. In type 2 diabetes, the body has insufficient production of insulin and impaired insulin sensitivity, and treatment consists of lifestyle changes with diet and exercise in combination, and medical treatment depending on need.

Individuals with diabetes have the burden of symptoms. They may have concerns about disease complications and may experience barriers in life due to the illness. Furthermore, they need to manage the disease every day of their lives. All these things may affect their quality of life negatively. Abilities to manage life with diabetes, e.g. being able to manage one's diet and stay physically active, with less diabetes complications and preserved good quality of life is a challenge and of great importance. A good ability may affect quality of life in a positive direction, possibly through greater confidence or through less worries and barriers. When an individual with diabetes acquires such abilities, I find it likely that this occurs with help and support from diabetes care.

The main risk factors for diabetes complications are well-known. We have looked at three important risk factors, namely glycated haemoglobin level (HbA1c), Systolic Blood Pressure (SBP), and Low-Density Lipids (LDL). They are important for diabetes complications and they are accessible and feasible to work with for diabetes care. They are also modifiable and therefore routinely measured and treated according to guidelines, to keep them within recommended intervals. There are other risk factors as well, but we focus on these three. Some typical complications are cardiovascular disease such as myocardial infarction and stroke (macrovascular complications), and microvascular diseases that cause damage to nerves, kidneys and the eyes.

The abilities, worries and barriers mentioned above can be measured with Patient-Reported Outcomes Measures (PROM). Experiences of diabetes care can be measured with Patient-Reported Experience Measures (PREM). PREM are likely to

reflect preferences, e.g. whether expectations of diabetes care are met. The three risk factors are easily obtained biomarker measurements, HbA1c and LDL are derived via blood tests, regularly taken, and SBP is measured using a blood pressure cuff.

Another burden associated with diabetes is that of healthcare costs. In Sweden, patients pay small co-payments, but there are schemes designed to protect individuals from high costs, e.g. cost caps for healthcare visits and caps for drug costs [2]. The remainder is tax-financed, and therefore the burden of healthcare costs falls mainly on society. These costs are substantial. In several studies in Sweden, diabetes was associated with high costs, e.g. a total of 5 billion SEK in Sweden in 1978 increasing to about 10 billion SEK in 2005 (expressed in year 2020 SEK) [3-6]. The cost levels varied by scope of costs and study methodology but were roughly in the same magnitude apart from a time trend. When separating total costs into costs for managing diabetes and controlling risk factors, and costs for managing complications, about a quarter of the costs were for managing/controlling, and about three quarters were for managing complications [4, 5]. In the most recent study, healthcare costs accounted for 37% of the costs, and costs of lost production due to permanent disability and mortality accounted for 63% [3].

Well-controlled risk factors should result in fewer diabetes complications, and consequently reduced healthcare costs for managing complications. At the same time, there are probably increased costs e.g. the drug treatment that resulted in well-controlled risk factors. Though the cost findings above may suggest overall cost reductions. Figure 1.1 illustrates a hypothesis of how diabetes care provides support, that affect abilities, risk factors and complications, and how diabetes care activities generate healthcare costs.

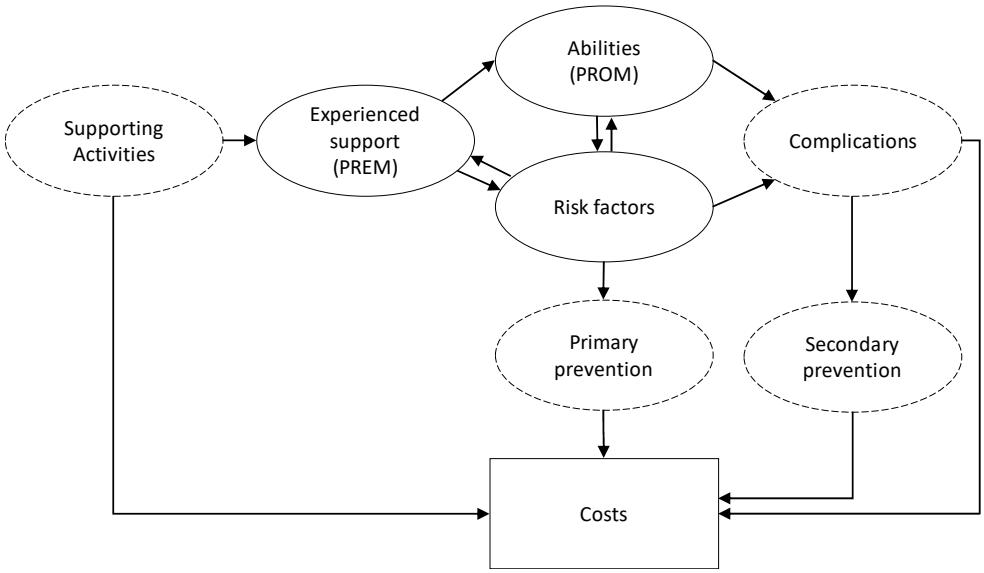


Figure 1.1: An illustration of my view of the process where diabetes care provides support (measured by PREM) that builds up abilities of managing diabetes and lifestyle factors (measured by PROM), and how this helps to control risk factors. This entails events and activities (dashed ellipses) resulting in costs.
Notes: Prevention refers to diabetes complications. There are bidirectional dependences since abilities, being self-rated, and judgment of support, e.g. satisfaction with treatment, may take impression of past risk factor levels.

How can we use PROM and PREM, and will they provide us with any important information that we cannot obtain already from the risk factors? According to the reasoning above, PROM would have impact on quality of life, and PREM too. We should therefore be able to measure quality of life using PROM and PREM. Quality of life would appear to be a very important goal for diabetes care.

Would it be possible to measure quality of diabetes care using PREM? If PREM reflect patient preferences, PREM ought to measure something like satisfaction. Is this in line with good diabetes care? Is satisfactory diabetes care good, or could it be poor? This depends on who defines good and poor, for instance whether preferences are weighed in into the decision or not. In addition, an individual might prefer short-term wellbeing to a good long-term prognosis. If it's difficult to achieve both, then the patient might aim for just one of them, i.e. to make a trade-off between them. Which one could also be a matter of preferences for the individual. This could result in poor care although experienced as satisfactory; or alternatively, good care experienced as unsatisfactory. Does the individual realize and understand the full consequences of the trade-off between short-term wellbeing and a good long-term prognosis? A more concrete question is whether PROM and PREM can be used for predicting future outcomes such as diabetes care costs and risk factors for diabetes complications. If high PREM predict lower future costs and risk factors, then we have support for the hypothesis that PREM reflect good diabetes care.

An individual perspective has surfaced here and there above. Abilities, worries, barriers, and experiences of diabetes care are subjective. They can be very individual in the sense that one individual's rating could differ from another's despite they live under objectively similar circumstances. Furthermore, a measurement of quality of life builds upon a view of what is important, and this is bound to vary from one individual to another. Preferences ought to be individual too. The individual perspective caught my attention, and I became interested in it as the theme of this thesis. It may not necessarily be of interest to everyone else. But there you have it.

This is a thesis in Public Health, within the subject of Health Economics. Health economics is a wide field. I touch only a few areas, estimation of quality of life, diabetes care costs, choice between alternatives (e.g. interventions, policies, approaches), and allocating resources. These areas are however important, when the need is great, resources are limited, and we wish that diabetes care can give individuals with diabetes a good life.

Aims

The overall purpose of the work described in this thesis was to promote an individual perspective in diabetes care through developing how patient-reported measurements are used in the evaluation of the situation of an individual with diabetes, how to work together with the individual in his or her situation, and in the improvement of diabetes care in general. This originated from a vision of the Swedish National Diabetes Register (NDR). We identified the following specific steps:

- As a first fundamental step to ensure that we could collect PROM and PREM, we wanted to define a procedure for establishing that PROM and PREM could be properly measured and capture relevant aspects.
- Next, we wanted to develop a method to take PROM, PREM and risk factors into account together, to measure overall quality of life, and identify causes of poor quality of life.
- Last, we wished to explore if PROM and PREM could predict future outcomes in terms of diabetes care costs and risk factors for diabetes complications.

This gave rise to a sequence of studies. The first two studies concerned development of patient-reported measurements, their measurement properties, and the role of these measurements as complements to traditional clinical outcomes measures [7, 8]. In all four studies, we explored ways in which these patient-reported measurements could be used to improve the situation of an individual with diabetes, as well as to develop diabetes care. We wished to explore how the measurements could be used to describe an individual's situation in terms of abilities, barriers,

opportunities, experiences of healthcare, and in the third study, as overall quality of life [9], and if these different measurements could be used to identify vulnerable individuals. In the fourth study, we wished to determine if patient-reported measurements could predict future outcomes in terms of diabetes care costs and risk factors for diabetes complications [10].

Motivation and potential impact

Regarding our aim to develop measurements of PROM and PREM. Provided that one agrees with the use of such measures, being able to obtain them must be fundamental. Secondly, being able to detect patients with low abilities, low judgments, or with high risk factor levels, one could trigger a response in the form of an intervention, and PROM and PREM could measure its effect.

A single measure of quality of life using biomarkers, PROM and PREM together would be useful for studying unmet needs and room for improvement in a holistic way, and for estimating the effects of interventions, and for comparing alternatives, in order to develop diabetes care with a broad focus on the individual.

The ability to predict future diabetes care costs and future risk factors for diabetes complication would be useful for detecting vulnerable individuals and help planning the allocation of resources, or perhaps even, help to proactively work to avoid future costs.

I hope our contributions may improve the situation of individuals with diabetes. Our vision is that an individual with diabetes can live as normal a life as possible while minimizing the risk of diabetes complications, and more generally to minimize the quality of life lost due to having diabetes. I envision that our work leads to useful tools for how healthcare and patients work together to accomplish as normal a life as possible without complications, and keeping worries and limitations imposed by diabetes to a minimum. Perhaps it can be useful also for setting priorities and allocating resources although this will be limited to within the field of diabetes only. In order to prioritize between different disease areas, generic measures of welfare (or health) are needed and an important limitation of my work is that it is diabetes-specific. This comes with the strength of being more sensitive to aspects important to a person with diabetes than a generic measure would be.

This thesis has been written in parallel with another thesis, concerned with the development of the instrument to capture patient-reported measurements [11], that we used in study II [8]. The two theses together make the foundation upon which any impact presented here rests.

A note on grammar and the use of first and third person: I wrote this thesis (though supervised) and any opinions and judgements expressed here are mine (see also acknowledgements). The thesis of course builds heavily on the studies I-IV, carried out by a collective of authors. Therefore, I describe what *we* found in these studies.

2. Materials and methods

2.1 Setting and data sources

This thesis consists of four studies carried out in Sweden. The study participants were adults with type 1 or type 2 diabetes registered in the NDR. The NDR has an estimated coverage rate of 94% of individuals with diabetes in Sweden [1].

Our study populations were samples of the registry population, subjected to questionnaire surveys to capture PROM and PREM as complements to their registry data of clinical variables in the NDR. In addition, we used data from other population-based registries, namely healthcare contacts from the Regional Claims Database VEGA in the Västra Götaland region, data on income and education from Statistics Sweden, and drug prescription data from the national prescription register. The use of personal identity numbers in Sweden allows linking of data from these registers on the individual level, for research purposes, provided ethical approval has been given.

As mentioned above, healthcare in Sweden is generally tax-financed with limited patient co-payments. The economic burden mainly falls on tax payors. Compared to many countries in Europe, Sweden has relatively high levels of public spending on health [2]. A national agency decides on the inclusion of new drugs in the health coverage policy. Apart from that, decisions on which methods, drugs, etc. to use, are taken at the regional or local level.

- In study I, the original NDR questionnaire was developed using two samples drawn from individuals in NDR living across Sweden. The main sample was used to develop scales, and a validation sample was used for validating the scale models. The scale scores used in the thesis came from the main sample.
- In study II, a new version of the NDR questionnaire was developed using a pilot survey of individuals living in the Västra Götaland region, and a larger survey of individuals living in the rest of Sweden. Scales were developed and scores were computed based on both samples combined.
- Study III was carried out on a subset of the full sample in study II, those with type 1 diabetes, with completely observed registry data and questionnaire data.

- In Study IV, a subset of the study population in study II was used, namely individuals living in the region of Västra Götaland, for which we could obtain data from the Regional Claims Database VEGA.

The resulting number of study participants, i.e. for study I and II, respondents to the questionnaires, are presented in Table 2.1 along with brief descriptive data.

Table 2.1: Summary of patient populations in studies I-IV (presented as number, mean (SD), percent).

	Study I. Main sample	***Study I. Validation sample***	**Study II**	**Study III**	**Study IV**
Type 1 diabetes					
Number of participants	1 124	*1 656*	1 849	1 456	468
Age	49 (15)	*50 (16)*	48 (16)	50 (16)	48 (16)
Duration of diabetes	24 (15)	*24 (15)*	25 (16)	26 (16)	25 (16)
Male/female	50%/50%	*50%/50%*	50%/50%	50%/50%	50%/50%
HbA1c, mmol/mol	63 (14)	*63 (13)*	61 (13)	61 (12)	60 (12)
Systolic blood pressure, mmHg	128 (15)	*127 (16)*	127 (14)	127 (14)	127 (15)
LDL cholesterol, mmol/l	2.5 (0.8)	*2.7 (0.7)*	2.5 (0.8)	2.4 (0.8)	2.5 (0.8)
Type 2 diabetes					
Number of participants	1 792	*1 431*	1 840	---	416
Age	66 (10)	*66 (12)*	66 (9)	---	66 (9)
Duration of diabetes	10 (8)	*12 (10)*	9 (8)	---	9 (7)
Male/female	56%/44%	*57%/43%*	61%/39%	---	62%/38%
HbA1c, mmol/mol	55 (14)	*58 (14)*	53 (12)	---	51 (12)
Systolic blood pressure, mmHg	137 (17)	*134 (16)*	134 (15)	---	134 (15)
LDL cholesterol, mmol/l	2.6 (0.9)	*2.6 (0.9)*	2.5 (0.9)	---	2.5 (0.9)

2.2 Ethics

The studies in this thesis were approved by the Regional Ethics Review Board in Gothenburg. This covered linking data from the NDR, the questionnaires, and other registers. Written consent was obtained from the respondents to the questionnaires.

2.3 Methods

This subsection describes the various methods used in this thesis. In study I, we identified a set of procedures to develop PROM and PREM scales and examine their measurement properties and compute scores on these scales. These procedures were then used in study II on a new version of the questionnaire. Section 2.3.1 describes these procedures. In study III, we examined if we could apply a method to estimate the quality of life, based on our data on PROM, PREM and risk factors (Section

2.3.2). Finally, in section 2.3.3, we describe how we tried to identify predictors of future costs and future risk factors, among our PROM and PREM.

In addition, specifically in study I, along with the first NDR questionnaire, we also used the generic health-related quality of life instrument EQ-5D as a known reference instrument. Using Swedish preference weights [12], we computed the EQ-5D-3L Index, serving to summarize the instrument into a scalar value of quality of life. We computed Spearman correlations between scales, risk factors and the ED-5D-3L Index.

2.3.1 Study I and II: Method for measuring PROM and PREM

Some properties can be objectively measured using instruments or laboratory tests, for instance, temperature (using a thermometer), body weight (scales), blood pressure (blood pressure cuff). This is not the case for every aspect that surrounds an individual with diabetes. In order to measure an individual's abilities, worries, barriers, or how they experience diabetes care, we need to ask the individual themselves. In this thesis, we used questionnaires, i.e. a set of questions to which the individual responds. We use here a methodology called Item Response Theory (IRT). In IRT, the term *item* is used instead of the term *question*. In order to translate responses (e.g. "I agree", "No", "Yes", "I disagree somewhat") into values that we can analyse, we develop scales and compute scores on these scales using a scale model. An example of a scale is the ability to manage diabetes. A scale is a group of questions (i.e. items) that address an underlying construct such as an ability, a judgment, etc. For this to work, data must satisfy a number of prerequisites, and this needs to be checked. The translation of responses into a value can be made provided the prerequisites are met.

If there is a deviation from the prerequisites, it needs to be taken care of. This could be to rephrase an item or split up a scale into two scales. If all prerequisites appear to be met, the questionnaire can be used, and measurements (PROM and PREM) can be computed from the questionnaire. At this point, we have not yet carried out all relevant checks, we have only assessed basic measurement properties. Developing a questionnaire is a long process, involving the construction of the items and carrying out different types of validation, and this has been described in another thesis [11]. Part of the validation work is still ongoing.

Addressing basic measurement properties still requires several steps to be taken. Some of these steps are quite technical. For instance, the questions in a scale should address only a single underlying construct, at least the scale models that we have used. Other aspects are more intuitive, and in my opinion more important for the reader of this thesis, and they are described as follows. For the specially interested reader, the methodology is described in further detail in study I (Appendix 1)

Differential Item Functioning (DIF) is a potential issue. This means that the item works in a certain way in one group of respondents that differs from how it works in another group of respondents. When we compute a score based on this item, we need to value it differently in the two groups. We then have difficulties comparing these two groups on the scale that uses the item. The best example of DIF that I have seen was in cancer. In the quality register for testicle cancer, a questionnaire is used to detect concerns related to manliness after removal of tumour-stricken testicles. Some respondents to the questionnaire have raised concerns regarding how this question might be used since their cancer was discovered during an investigation for a sex change [13]. For them, lost manliness was no concern at all, but rather the whole point. In terms of DIF, this item indicates a negative impact for some individuals, and a positive impact for other individuals (assuming a binary view of gender), which an analyst needs to handle. Admittedly, the sex change group may not be very large. But in this example as well as in general, DIF needs to be considered, whether to make an adjustment or not, whether the issue is important (as individuals might certainly argue), and whether it would impact on the purpose of the questionnaire. This example may not be central in diabetes, but it describes important aspects from the perspective of some respondents, and how that leads to the occurrence of DIF.

Two aspects are related to the fit of the scale model. We need to examine every item and check that its response data fit the model, so-called *item fit*. In addition, the *overall model fit* is checked for each scale model. Good model fit is important for the confidence in a scale model and the scores we compute using the scale model.

There are no absolute criteria for judging if the various prerequisites are met. E.g. one sometimes says that there is always DIF, but the important question is if the DIF has any practical impact. Large sample sizes increase the statistical power to detect deviations, and one might need to make a partially subjective judgment whether any deviations are important enough to reject a scale (or a scale model).

There are also other benefits with IRT. If scales successfully pass the review described above, they allow us to estimate latent constructs like PROM and PREM. They hereby reduce a multidimensional set of items into a single estimate. Although the items each one by itself carries interesting information, it is more practical to have a single value summarizing the items and their responses. Furthermore, we get a measurement of the precision of the estimate, and the estimate is robust to partially missing data (even if the individual didn't respond to some items, we can use the scale model with the other items to compute a score). IRT assigns to each item a "difficulty". Difficulty may vary between items i.e. different levels of the underlying construct are required to endorse a given response level of the items. This makes items able to measure the construct at different segments of the scale, e.g. one item

is good for measuring low levels of the construct while another item is good for measuring high levels of the construct.

We carried out a validation of the scale models, using a separate validation sample. We fitted new scale models using the validation sample and used them to compute alternative scale scores from the response data in main sample. The original scores were then compared to the alternative scores, to study how robust the methodology was able to fit scale models.

An important aspect is what a specific value on a scale means. What does it mean if my freedom from worries is 75 on a scale ranging from 0 to 100? Ideally, we would have reference values that everyone understands and agrees on.

2.3.2 Study III: Method for measuring the quality of life

There are several ways of measuring the quality of life. Perhaps the most commonly used method is to use a health-related quality of life questionnaire together with a value set, for instance, the EuroQol (EQ-5D) [14] along with a Swedish value set [12]. Another method is to elicit individuals' utility functions directly or to carry out a willingness to pay study [9]. Still, other options are to go through time trade-off or standard gamble-exercises [15]. We saw difficulties with these methods that inspired us to look at efficiency analysis instead, a method that could perhaps resolve some of the difficulties or complement them.

We decided to examine in study III if efficiency analysis could be used [16]. Efficiency analysis builds on the idea of production of some output, and the production requires raw materials that we call input. There can be several inputs and several outputs. The units carrying out the productions, in our case individuals with type 1 diabetes, need not produce the same amounts of outputs, or use the same amounts of inputs, as any other individual. They can use their own mix of inputs to produce their own mix of outputs. For any individual, we think of inputs and outputs as being weighted according to their importance for the specific individual. The method does not require any specific set of importance weights, nor does it require the same set of weights to be used by everyone. It compares individuals with similar weights to each other and results in a single measure of relative accomplishment, so-called *efficiency*, assumed to be comparable across different sets of weights [9]. The freedom to choose one's own weights is called *Benefit of the Doubt* in the literature [17]. Though we grant freedom in the specific weights, we do relate an individual's amounts of outputs produced to its consumed amounts of inputs. We consider the production of more output, and use of fewer inputs, beneficial. The limit of what is possible to produce given the inputs is called *technology* in efficiency analysis. Technology can change - e.g. through inventions, new drugs, or new approaches.

Our production model of quality of life was inspired by the capability approach [18], a framework that we partially implemented. Our model consists of two parts, an intermediate model that uses aspects of support from diabetes care to produce the ability to manage diabetes and the lifestyle factors diet and physical exercise (Figure 2.1). In a second part, the capability model, these two abilities are used along with freedom from worries and barriers, to produce wellbeing and well-controlled risk factors. We designed this model both from the capability approach and from our view of diabetes care illustrated in Figure 1.1.

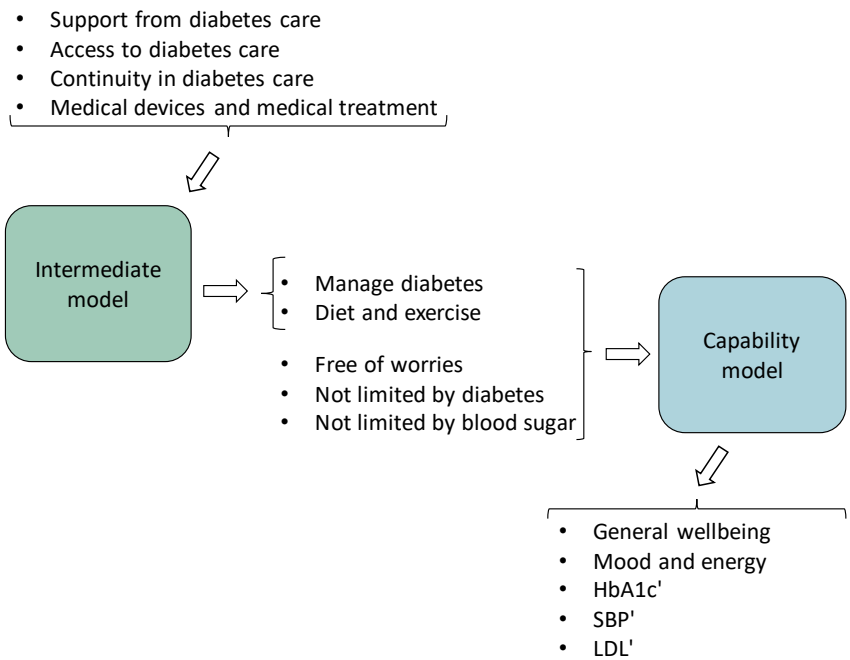


Figure 2.1: The production model for quality of life.
Notes: the risk factors are transformed (') to work as outputs, namely well-controlled risk factors.

We used Data Envelopment Analysis (DEA) to operationalize the efficiency analysis [16]. The analysis method models the frontier, which is the set of most successful individuals, namely those who according to their inputs produced the most outputs. The frontier is taken as the technology, that is, what is possible to produce. Allowing individuals to use their own mix of outputs means that the frontier consists of different segments with different mixes of outputs. For instance,

- Very well-controlled risk factors but not so good wellbeing.
- Poorly controlled risk factors but good wellbeing.
- Well-controlled HbA1c and fair Mood and energy, but poorly controlled SBP and LDL and poor General wellbeing.
- And so on, many imaginable mixes of these five outputs.

Every individual is compared to other individuals with the same mix, and this means that the individual is judged relative to the segment on the front where there is the same mix of outputs. Figure 2.2 illustrates a simplified example. In the example, Carol is compared to Adam since they both make the same type of trade-off between wellbeing and well-controlled HbA1c. Their focus is on wellbeing. On the other hand, Bella and David focus on HbA1c and accept poorer wellbeing. There is no correct answer as to the mix, it comes down to an individual's preferences (although it could also be considered a normative question).

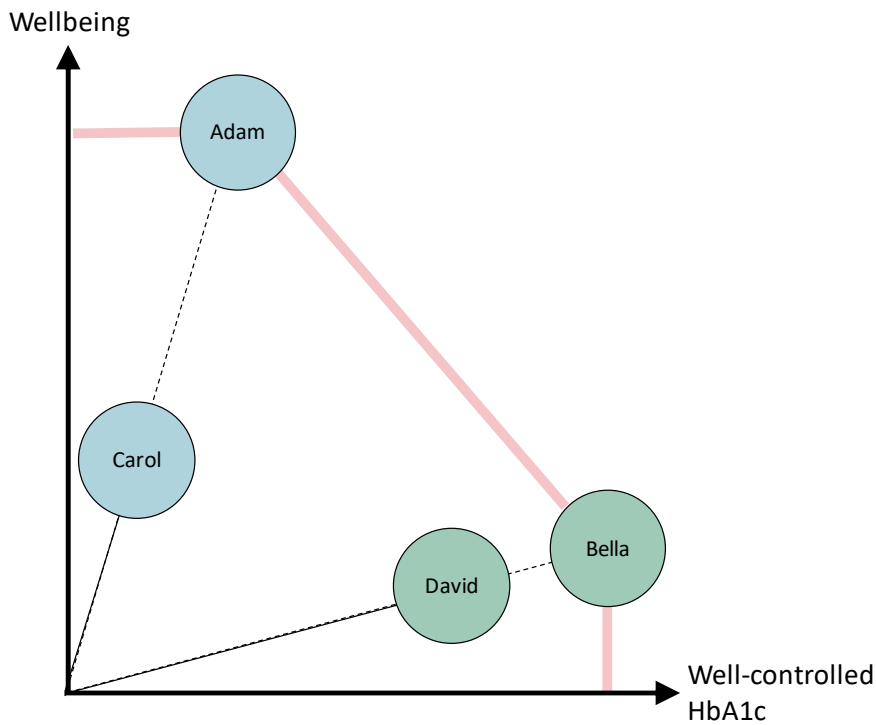


Figure 2.2: Simplified efficiency-analysis example: The frontied (pink line) represents a trade-off between wellbeing and well-controlled HbA1c. Adam and Bella are efficient (successful), and define the frontier. Carol (inefficient) is compared to Adam. David (inefficient, but more efficient than Carol) is compared to Bella. The black lines show Carol's and David's efficiency estimates (solid black) and their projections onto the front (dashed black).The dashed black lines also indicate their output shortfall.
Note: the names of my imaginary friends have been assumed, to avoid disclosure.

After conducting the efficiency analysis, we obtain for every individual, an estimate of success (efficiency), and estimates of output *shortfall*, i.e. the amount of output *not produced* by the individual. In the figure, the solid line to David represents David's output, and the dotted line from David to Bella represents David's output shortfall. If David were fully successful like Bella (i.e. efficient), he would have produced more output, and he would then have been located at the same place as Bella in the diagram. Please note that the diagram only illustrates shortfall with respect to inefficiency. Another reason for shortfall is having less inputs (see paper III for more detail). To summarize, output shortfall can be decomposed into a part due to inefficiency, and a part due to the level of inputs that the individual had access to (compared to some reference level). We also obtain a measure of contribution to the efficiency, from each of the outputs. I.e. how much *General wellbeing* contributed, how much *Mood and energy* contributed, etc, to the individual's efficiency, or in our application, quality of life.

2.3.3 Study IV: Method for finding predictors of future outcomes

We used regression analysis to detect predictors of future outcomes. PROM and PREM were used as explanatory variables in regression models of future diabetes care costs and future risk factors (one at a time). This would reveal associations between the PROM and PREM, and the future outcome. Since the data on the outcome is from a later point in time, this means we may detect predictors in a forecasting sense. It may even lend a bit of support to a hypothesis of causation, e.g. that high PROM results in a lower cost. But here one must thread with care. Even though the PROM occurs first, and the cost occurs later, there could still be a third factor that explains both their values and this could be misinterpreted as PROM causing the cost. By controlling for other factors that could likely be such third factors, we reduce the risk of making this type of mistake. However, we still cannot prove causation. We, therefore, look for predictors to be on the safe side. We had data on patient characteristics such as age, gender, duration of diabetes, risk factors, costs, presence of other chronic illnesses, education level and income, from before the questionnaire was filled in, and we used these data to control our models.

Costs were categorized as inpatient costs, outpatient costs, and drug costs. We carried out a regression analysis on each cost category as well as on total costs, one at a time. Total costs also included costs for insulin pumps and medical devices. Then we carried out a regression analysis on each of the risk factors HbA1c, SBP and LDL.

3. Results

Here I briefly summarize the results of studies I-IV, and what they accomplished. Please note that study I resulted in one set of scales, and study II resulted in another set of scales, the latter subsequently being used in study III and IV. Unfortunately, the scale names are not always the same in manuscript IV (Table 3.2). For further details, please see the separate papers in appendix 1-4.

3.1 Study I

This study contributed by identifying and trying out a set of procedures for developing IRT scales and determining measurement properties of the scales based on the first NDR questionnaire. The same set of procedures were used in Study II with the new version of the NDR questionnaire that was developed later using a more patient-centred approach, and this may be considered its contribution from the perspective of the thesis. However, study I also contributed with important findings based on the first questionnaire, and these are described below. We also have some ideas for follow-up research (see Future perspectives).

We defined scales based on the themes covered by the questionnaire. One of them, diabetes self-management, had to be split into two separate scales because there were two underlying constructs. In total this resulted in eight scales that were found to have acceptable measurement properties. There were five PROM scales comprising diabetes self-management abilities (one scale for knowledge and skills, one for stress- and satisfaction-related items), sense of security, and abilities to participate in work activities, and in social activities. There were also three PREM scales (though paper I used the term PROM for all scales) covering access to healthcare, service and information, and involvement in decisions regarding the patient's treatment and care. Scale models were fitted for these eight scales, one set for type 1 diabetes, and another set for type 2 diabetes. Using a separate validation sample, we validated the scale models with very satisfactory results.

When we looked at our sample in terms of scale scores, we found that younger individuals with type 1 diabetes had lower scores of diabetes self-management and judged Service & information and Access lower than older individuals (split by

median age). We saw this in type 2 diabetes as well, and that younger individuals had lower scores on sense of security. Neither gender nor diabetes duration had however any influence on any of the IRT scores.

As a way of examining if the scales could be used as signals for individuals that needed interventions from diabetes care, we looked at subgroups with very low scores on each of the scales. We made a few findings, e.g. that low self-management ability was associated with higher HbA1c in type 1 diabetes, and that low sense of security was associated with higher HbA1c in type 2 diabetes. We also looked at subgroups of individuals that had high risk factor levels, and there were corresponding patterns in individuals with high HbA1c: e.g. having low Self-management ability and a low sense of security (in both type 1 and 2 diabetes). See section 4.3 in Paper I for further details.

The NDR questionnaire was issued together with EQ-5D, from which we computed the EQ-5D Index. We could study the correlation between our scales, risk factors for diabetes complications, and the EQ-5D Index. In type 1 diabetes, the strongest correlation with EQ-5D Index was seen in two PREM-scales, Service & information and Involvement, and the PROM scales Sense of security and Self-management ability are almost as strongly correlated (Table 3.1). In type 2-diabetes, we saw a very similar pattern, though we also see correlations among some of the other scales. The only risk factor significantly correlated with EQ-5D Index was HbA1c, and only in type 1 diabetes. It was negatively correlated (Table 3.1).

Table 3.1: Correlations (Spearman's rho) between latent variables, risk factors, and EQ-5D Index.

	Type 1 diabetes EQ-5D Index	Type 2 diabetes EQ-5D Index
PROM		
Self-management skills	0.21 *	0.24 *
Self-management ability	0.34 *	0.32 *
Sense of security	0.37 *	0.31 *
Social activities	0.06	0.24 *
Work activities	0.14 *	0.22 *
PREM		
Access	0.13 *	0.20 *
Service & infomation	0.42 *	0.42 *
Involvement	0.38 *	0.49 *
Risk factors		
HbA1c	-0.13 *	-0.06
LDL	-0.04	0.04
SBP	-0.09	0.06

Notes: Correlations computed as Spearman's rho. * = significant correlation, $p < 0.0001$.

In study I, we concluded that we had managed to collect data to estimate patient-reported outcome measures in the form of patient abilities and judgments of their experience of diabetes care. Together with risk factors for diabetes complications, they describe different aspects of a patient's situation. These aspects occasionally overlap, but not in any particularly useful way. They all provide important for decision-makers and none is necessarily more relevant than the other, for a more complete evaluation of diabetes care and for promoting person-centred care.

3.2 Study II

In paper II, we used an updated version of the questionnaire, that reflected aspects important to individuals with diabetes, identified by individuals with diabetes, and phrased using the individuals' own words. This resulted in a more person-centred questionnaire than the first questionnaire. We obtained scales with satisfactory measurement properties, better than those based on the first version of the questionnaire. There were PROM scales addressing wellbeing (general wellbeing, mood and energy), freedom from worries and barriers (Free of worries, Not limited by diabetes, Not limited by blood sugar), and support by family and friends (Support from others), and PREM scales Support from diabetes care, Access to diabetes care, Continuity in diabetes care and satisfaction with medical devices and medical treatment (Table 3.2). All scales had acceptable test-retest reliability, and many could detect differences between diabetes types, age gender and treatment subgroups.

In several aspects, e. g. Free of worries, type 1 patients reported lower scores than type 2, and younger patients reported lower scores than older in both diabetes types (Figure 3.1).

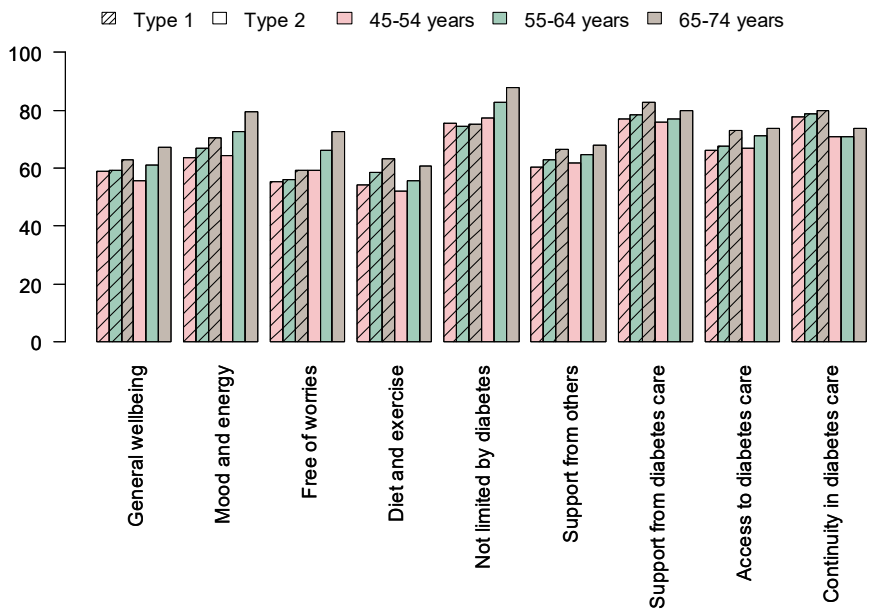


Figure 3.1: Mean IRT scores by age stratum in type 1 and type 2 diabetes (Paper II).

As treatment was stepped up from diet only, to oral medication, and on to insulin treatment, individuals with type 2 diabetes showed subsequently lower scores on several scales (Figure 3.2).

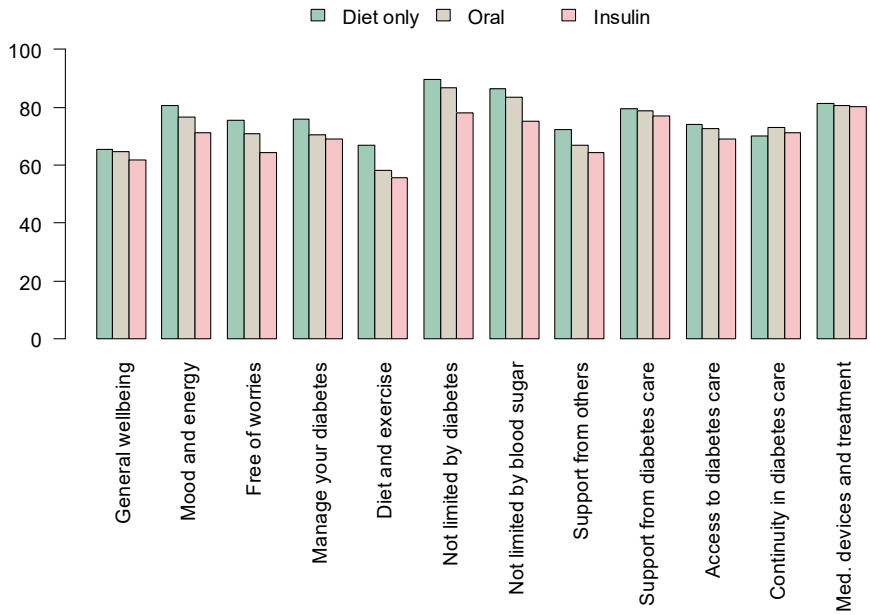


Figure 3.2: Mean IRT scores by treatment group in type 2 diabetes (Paper II).

HbA1c, SBP and LDL-cholesterol (LDL) consistently showed low correlations with each other as well as with IRT scores (Study II; online supplementary table S4). In type 1 diabetes, the strongest of these low correlations was between HbA1c and ManD (−0.23, $p<0.001$) and in type 2 diabetes, the strongest correlation was between HbA1c and FreW (−0.25, $p<0.001$). None of the scales was significantly correlated with any of the other examined risk factors, i.e. SBP and LDL.

We also looked at vulnerable individuals, as a means to identify individuals in need of an intervention. Hereby we saw that in type 1 diabetes, low Manage diabetes scores were associated with higher HbA1c levels than type 1 overall, and vice versa. There were similar patterns in type 2 diabetes, though involving more scales: low scores in Manage diabetes, freedom from worries and barriers, as well as Mood and energy all corresponded to high HbA1c and vice versa.

Looking from the diabetes care perspective, clinicians reported a positive experience of using scores, visually presented, in the patient dialogue. This was explored in a handful of pilot clinics across Sweden, where patients filled in the questionnaire immediately before their visit, and the scores were available for discussion during the visit (data not shown).

Based on study II, we concluded that the new questionnaire with items phrased in accordance with the patients' own words, could collect data and estimate patient-reported outcome and experience measures in the form of well-being, abilities and judgments of diabetes care. We had hereby taken a new step towards a broader evaluation of diabetes care and more person-centred care. The measured well-being, abilities and judgments of diabetes care appeared to comprise a useful complement to risk factors for diabetes complications, and they reflected several aspects of patient-experienced living with diabetes and diabetes care. We could also identify where these aspects could be improved.

Table 3.2: Scales based on the second NDR questionnaire (study II).

Scale	Abbreviation used in papers II and III	Scale name used in manuscript IV
PROM		
General wellbeing	GenW	General wellbeing
Mood and energy	MoE	(not used)
Free of Worries	FreW	Free of worries
Manage your diabetes§	ManD	Manage diabetes
Diet and exercise	DiEx	Diet and exercise
Not Limited by Diabetes	NLD	Not limited by diabetes
Not Limited by Blood Sugar§	NLBS	(not used)
Support from Others	SuO	(not used)
PREM		
Support from Diabetes Care	SuDC	Support in diabetes care
Access to Diabetes Care	AcDC	Access in diabetes care
Continuity in Diabetes Care	CoDC	Continuity in diabetes care
Medical Devices and Medical Treatment§	MDMT	Satisfaction with treatment

Note: § separate scales for type 1 and type 2 diabetes.

3.3 Study III

In the third study, we developed a way to use PROM, PREM and risk factors for diabetes complications together to evaluate the situation of a person with type 1 diabetes in terms of quality of life. We reviewed a methodology, efficiency analysis using Data Envelopment Analysis, and judged that it was adequate in our application. Hereby, measures were developed, that described how efficiently diabetes care and the person together create abilities and wellbeing. Please see the Methods for details on inputs, outputs and efficiency. Our approach comprised a sequence of two models. an intermediate model and a capability model.

One desired aspect of our measurement of quality of life was that it would discriminate between different individuals and therefore be useful for comparisons. We used several variables to construct our measure, and multiple comparisons could lead to ambiguous rankings. However, the measure discriminated between most

individuals, in the intermediate model between all but 4% of the individuals and in the capability between all but 9% of the individuals. Thus, only a minority were indistinguishable from each other in terms of quality of life.

Further, the work aimed at identifying what aspects could be improved for a given individual and how much they could be improved. Every individual's set of outputs was judged in relation to its projection onto the frontier (and inputs were taken into account as well), and the individual's output shortfall, i.e. the output they did not express was hereby determined. Table 3.2 shows three examples.

- Individual 1 fully successfully produces quality of life (i.e. is efficient), located on the frontier and therefore efficient, and has no output shortfall. At the segment of the frontier where this individual is located, the PROM scales are at their maximum, but the risk factors are only moderately well-controlled.
- Individuals 2 is less successful (i.e. is inefficient) and lies below its projection on the frontier and therefore has non-zero output shortfall. The individual lies roughly in the middle of the scale on all five outputs and has similar shortfalls across the five outputs. It belongs to a frontier segment with only moderately high levels on all five outputs.
- Individual 3 is inefficient too but has relatively high *Mood and energy* and HbA1c but low *General wellbeing* and SBP. Its highest output shortfall is on *Mood and energy* and HbA1c, because the frontier segment on which it is projected has almost perfect *Mood and energy* and HbA1c.

Table 3.2: Examples of individuals with observed outputs, and output shortfall due to inefficiency.

	Individual 1	Individual 2	Individual 3
Observed outputs			
General wellbeing	100	58	37
Mood and energy	100	59	69
HbA1c (transformed)	74	55	71
SBP (transformed)	63	50	36
LDL (transformed)	49	63	50
Efficiency (Quality of life)	1.00	0.73	0.72
Outputs at frontier			
General wellbeing	100	80	51
Mood and energy	100	81	96
HbA1c (transformed)	74	75	99
SBP (transformed)	63	68	50
LDL (transformed)	49	87	70
Output shorfall			
General wellbeing	0	22	14
Mood and energy	0	22	27
HbA1c (transformed)	0	21	28
SBP (transformed)	0	19	14
LDL (transformed)	0	24	20

Notes: Risk factors transformed to act as outputs, so that higher values are more beneficial.

We also studied a set of clinics, to learn why output shortfall occurred. Here we looked at intermediate production at the clinic level (average of their patients), and separated output shortfall into a part due to inefficiency, and a part due to having fewer inputs, i.e. the level of support from diabetes care as judged by the patients, compared to that at a reference clinic. We saw that output shortfall was mainly due to inefficiency (Figure 3.3).

The efficiency models also revealed how much each output contributed to total efficiency, i.e. quality of life. We saw that HbA1c was overall the strongest contributor, but this varied between the least and the most efficient individuals. Figure 3.4 shows how the risk factors are the strongest contributors in the least efficient quartile group, especially HbA1c. This changes gradually with efficiency, and in the most efficient quartile group, the PROM scale *Mood and energy* is the strongest contributor, and HbA1c is the third strongest. These are however group averages (see also Appendix 3, Table 4), and examining this in further detail reveals fragmented patterns, i.e. considerable individual variation (data not shown).

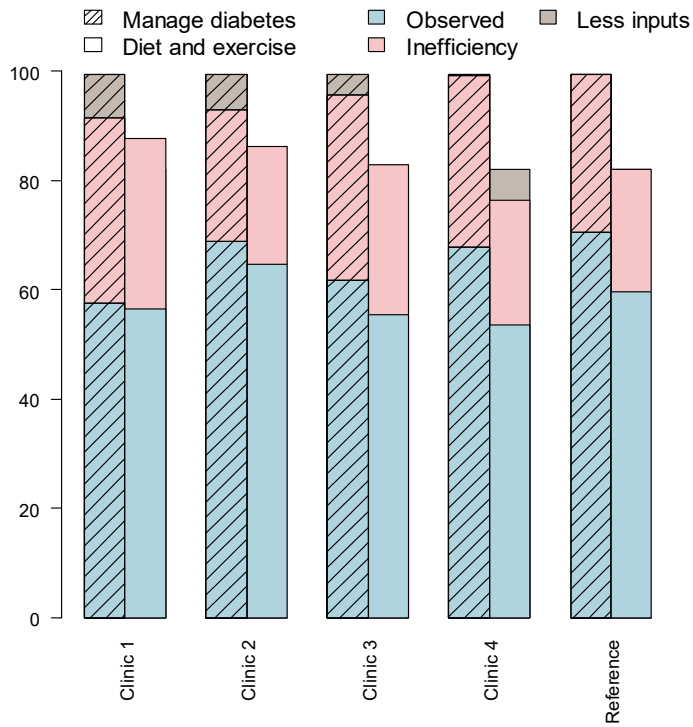


Figure 3.3 : Mean outputs (observed) at five diabetes clinics, mean inefficiency-related output shortfall (Inefficiency), and mean input-related output shortfall (Less inputs) compared to if clinics 1–4 had the same input levels as the reference clinic.

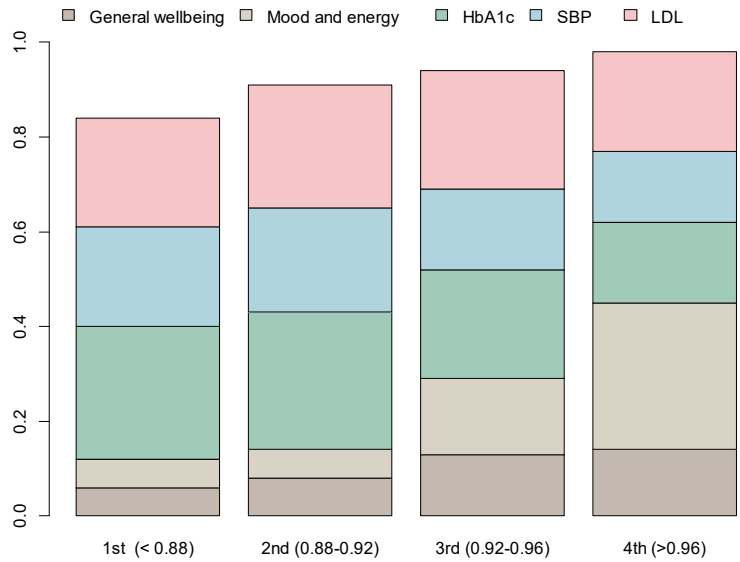


Figure 3.4: Contributions to quality of life (Capability efficiency) from PROM scales and risk factors, in the quartile groups (1st = least efficient, 4th = most efficient).
Notes: The risk factors were transformed so that a well-controlled risk factor contributed more than a poorly controlled. The range of efficiency is shown withiin parenthesis).

To summarize study III, the efficiency analysis approach could use PROM, PREM and risk factors to estimate the quality of life with a broad focus on the individual, in individuals with type 1 diabetes. The approach enabled ranking and comparisons using all these aspects in parallel and allowed every individual to express their own view of which aspects were important to them. We judged that the approach could be used for policy regarding interventions on inefficiency as well as healthcare resource allocation, although currently limited to type 1 diabetes.

3.4 Study IV

In our fourth study, our main question was to study if a set of PROM and PREM scales predicted future diabetes care costs and future risk factors for diabetes complications. We also examined if their effects on future costs and risk factors were influenced by controlling for costs and risk factors during the year preceding the questionnaire that we used to collect the PROM and PREM.

In type 1 diabetes, we found no PROM nor any PREM with any effect on future costs that was consistent over model specifications, with one exception. High scores on *General wellbeing* had a negative effect on future inpatient care costs. The effect

corresponds to about 2% reduction of inpatient costs per unit General wellbeing, or almost 90% reduction in case General wellbeing increases from its theoretical minimum to its maximum. In type 2 diabetes, we did not detect any effect on future costs from any PROM nor any PREM, and this was consistent across model specifications, and across different cost categories.

When we looked at predicting future risk factors, we found that in patients with type 1 diabetes, high *Manage diabetes* predicted low future HbA1c, consistently across different model specifications. In patients with type 2 diabetes, high *Satisfaction with treatment* predicted low future HbA1c and this was also consistent across different model specifications. Neither PROM nor PREM scales appear to have any consistent effect on the risk factors SBP and LDL, neither in type 1 nor in type 2 diabetes.

The effect of *General wellbeing* on inpatient costs was not influenced by controlling for costs and risk factors during the preceding year, so we do not need to control for them. The effects of *Manage diabetes* and *Satisfaction with treatment* on HbA1c were, however, tripled and doubled, respectively when we didn't control for HbA1c during the previous year, so here we need to control for previous HbA1c. On the other hand, their effects were not influenced by controlling for previous costs so controlling for them is unnecessary.

Among the factors that we controlled for in the study, duration of diabetes and the presence of other chronic diseases predicted costs in some cases. Duration and renal failure predicted future total costs in type 1 diabetes (manuscript IV, Supplementary Table S1). Looking at the cost categories, duration mainly affected the outpatient costs, whereas renal failure affected inpatient costs. In type 2 diabetes, duration predicted higher total costs, as did mental health problems and cardiovascular disease (manuscript IV; Supplementary Table S8). Mental health problems predicted inpatient costs, and duration och cardiovascular disease predicted drug costs (See Discussion).

4. Discussion

4.1 Measuring PROM and PREM, and their relation to risk factors

We examined how others had used IRT to assess the measurement properties of their instruments and we tried to collect a suitable set of procedures to make the necessary investigations and checks and applied it to the first NDR questionnaire. This resulted in study I, where most of the methodology was taken from published literature. We managed to assess the measurement properties of scales in the first questionnaire and found them acceptable. Having access to a second sample, we were able to validate the set of procedures and our resulting scale models, with good results. This speaks well of the scales, but also (to my judgment) of our approach.

In study II, we applied the same set of procedures to assess the new version of the NDR questionnaire, whose measurement properties were found to be somewhat better than those of the first questionnaire. Most of the scales were common to type 1 and type 2 diabetes, which was an improvement over study I. Common scale models allow us to make comparisons between type 1 and type 2 diabetes. Another improvement was that the new questionnaire was constructed by consulting individuals with diabetes to identify which aspects they found important [19], however, this was not part of the current thesis. The questionnaire in study I was constructed based on expert opinion.

Both study I and study II had considerable strength in their large sample sizes for developing IRT scales, and that we had data on risk factors linked on the individual level. These data allowed us to study correlations between scales and risk factors.

A limitation of both the old questionnaire in study I and the new questionnaire in study II is that we lack reference values. This is a consequence of developing a new questionnaire, and reference values need to be derived through future research. When the first questionnaire was developed - though this was before I joined the project - there was no existing questionnaire that covered all the desired aspects, so it was deemed necessary to develop a new questionnaire. Another inevitable consequence of a new questionnaire is the lack of international comparisons - and a questionnaire in Swedish does little to help. But the questionnaire can be translated, and the NDR has received inquiries from other countries to do so. On the other hand,

these questionnaires are very relevant for the Swedish diabetes population, especially the new NDR questionnaire. They are however diabetes-specific which can be considered both a limitation and a strength since the aspects covered by a disease-specific questionnaire may be more relevant.

In both study I and study II, we examined how PROM, PREM and risk factors correlated. Of course, PROM, PREM and risk factors all measure different things from a theoretical point of view. But we saw only occasional correlations between PROM, PREM and risk factors. Our conclusion was that they were complements to each other. There is solid evidence that risk factors provide important information, and I mention Cederholm et al, Zethelius et al, and Rawshani just to provide a couple of examples [20-22]. Based on ideas of reasonable determinants of quality of life and preferences, we have argued that the PROM and PREM in our questionnaires are important as well [7, 8]. Further to this, we found in study III that we could measure the quality of life based on all three categories of data [9]. So, I persist in claiming that these three categories of data are all relevant for understanding the situation of an individual with diabetes.

We also looked at concrete ways of using the data. Subgroups with IRT scores below the tenth percentile, and subgroups defined as having a risk factor level above the 90th percentile were examined om study I and II if the subgroups differed from the full sample with regards to other values as well. Here we made occasional findings, though the overall impression remains that PROM, PREM and risk factors are important complements to each other.

4.2 Measuring quality of life

After reviewing the methodology for our adaptation, we used efficiency analysis to estimate the quality of life from PROM, PREM and risk factors. Our model in two parts used PREM as inputs to produce outputs, namely the ability to manage diabetes and the ability to manage diet and exercise. These abilities, along with freedom from barriers and worries, were used as inputs into the production of general wellbeing and mood and energy, and well-controlled HbA1c, SBP and LDL, as outputs. We took impression from the capability approach when we designed this two-part model [18]. So, we were indeed not the first to think along these lines. Månsdotter et al suggested a somewhat similar approach for a generic measure for monitoring and evaluating social welfare and public health [23]. They discussed their suggested approach from the perspectives of different normative theories. Further, they considered which capabilities would be important to include. We concentrated on operationalizing an idea, starting with the very specific situation of diabetes, and we wanted to use aspects that individuals with diabetes themselves

had identified as important. We added a set of well-known and important risk factors for diabetes complications, in order to incorporate future health.

Furthermore, we chose a method that used weights sensitive to an individual's valuation of importance. Perhaps a better way to express this is that the weights were derived from which outputs the individual expressed and that we assume the individual expressed these outputs according to what was possible for the individual and the individual's preferences. We saw that there was great heterogeneity in how much different outputs contributed to an individual's quality of life (Figure 3.4). So, allowing individual weights apparently had a beneficial impact on the analysis. One could speculate that not allowing individual weights would force all individuals to comply with the most common set of weights, and this would likely have made much fewer individuals be among the successful. It would be less person-centred, but whether that is a disadvantage is a normative question.

One might question our approach of using efficiency analysis - and indeed some people have. But we believed that our approach could be useful, and we persisted and described our first steps and discussed their merits and limitations. Perhaps someone else can benefit from our attempt. However, we do not underestimate the amount of further work required to validate various remaining aspects of our measure (see Future Perspectives). Or, as Månsdotter et al put it, "Of course, years of empirical research and refinement remain" [23]. This probably applies to our work as well. We concluded, however, that our approach worked. The levels of outputs were consistent with production efficiency, i.e. the more efficient, the higher the outputs. Ideally, we would have been able to compare our estimates with some other already established estimate of the quality of life. Unfortunately, we did not have such data (though see Future perspectives). In study I, we had the scales from the first questionnaire as well as EQ-5D, from which we derived the EQ-5D Index. We saw some correlations between our scales and EQ-5D. Although these were not strong, they indicated at least some support for our presumption that the scales were connected to determinants of quality of life. This set of scales did however not include general wellbeing, like the set in Study II. Perhaps general wellbeing would be more strongly correlated with EQ-5D Index. Unfortunately, we did not use EQ-5D in study II. But a comparison with EQ-5D would only show the level of similarity between the two. An improvement over the EQ-5D Index (if one were so bold as to assume this possibility) could never be demonstrated by comparing to EQ-5D Index as it would result in lack of similarity in the comparison. But being worse than EQ-5D would also result in a lack of similarity. We need to acknowledge that we cannot differentiate between these two possibilities.

Universal reference values for our measure would have been useful. For instance, the widely used measure Quality-Adjusted Life Year (QALY), is a quality of life measure based so-called QALY weights. The weights are defined using two

reference levels, 1=Full health and 0=Being dead [15]. Our study IV could have been useful for determining reference values for the quality of life measure. If the variables upon which the measure was based had proved to be predictors of costs, we would have had a translation between the measure and costs, i.e. some evidence of value. Unfortunately, we detected very few predictors of future costs among the PROM and none among PREM, and the risk factors being predictors therefore only provide partial information.

What about our production model? Could one also think of some of the PREM measures as technology factors, things that catalyse quality of life? From the individual's immediate point of view, they may not be able to change the aspects of support that PREM measures. Thus, these could be considered non-discretional inputs, or technology factors, and be treated as such in the analysis. But we look at this in a wider context, that developing diabetes care can affect the PREM inputs (presumably an improved aspect is given a more beneficial judgment). So ideally, by reporting low PREM the individuals will drive the development of diabetes care and hence increase the inputs available. For this reason, we kept the PREM as inputs in the analysis. On the other hand, in study IV we saw no signs of relationships between PREM and costs, and we probably would, if PREM and resource use were strongly linked. Perhaps the PREM reflect what goes on in the dialogue between patient and healthcare staff rather than the amount of time, procedures etc. This would indicate that they are catalysts rather than inputs. In retrospect, after carrying out study IV, perhaps we ought to reconsider the roles of the PREM. Another aspect pertaining to our model specification is that we use both *Satisfaction with treatment* and *Manage diabetes* as inputs to the production of several outputs of which (well-controlled) HbA1c was one. We found in study IV that these scales predicted future HbA1c. This lends at least some support to our model specification in study III.

Another aspect connected to PREM is that being a judgment of diabetes care, it may be complicated to use individual PREM values in dialogue with healthcare staff. For instance, it could be uncomfortable for a patient to discuss a PREM score reflecting a lack of satisfaction with diabetes care. Perhaps PREM had better be presented only on the group level. Would our measure of the quality of life still work if the PREM involved were aggregated on a group level? (E.g. the average scores among patients visiting a given diabetes care provider). We have not investigated this, but we would in such a case deviate from the individual perspective. The judgments are individual, and one might wonder if it would be sensible at all to aggregate them over a group of individuals.

What would have happened, if we had used the scales from the first questionnaire (study I) to measure the quality of life? The set of scales in study I is different from the set of scales in study II, although some similar scales can be found in both sets. For instance, the scale *Sense of security* in study I might be similar to *Free of worries*

in study II, and both sets have scales for the ability to manage diabetes and access to diabetes care. But there is another more subtle difference. The items in study I were developed by experts or picked by experts from other questionnaires. In study II, the items were derived from interviews with individuals with diabetes, to include aspects they thought were important. Therefore, the scales in study I may be considered less person-centred and this could result in a different valuation of the scales. It is difficult to predict the impact of this on the quality of life score, but in my opinion, aspects identified with the help of individuals with diabetes fits very nicely with the method that allows every individual to use their own importance weights on the aspects included in the calculation. Apart from these differences, the scales in study I would meet the prerequisites of the method no less than the scales in study II. Since we used EQ-5D in study I, it is tempting to think more about this (see Future Perspectives).

How could we use our efficiency measurements to improve the situation of individuals with diabetes? We see three main strategies. First, we could try to make the individuals become more successful (efficient), by themselves. What if we could somehow inspire them to be more like more successful individuals, that share their view of what is important (i.e. having the same mix of outputs)? Perhaps by showing them what is possible. Though I lack credible ideas as to how we specifically would do this. Secondly, we could intervene on the causes of their inefficiency. Imagine that efficiency comes partly from diabetes care (and its support). We could look at caregivers that have successful patients, try to learn from how they work. If we discover what they do that makes their patients successful, we could prompt other caregivers to work in the same way. Another approach would be to exploratively regress efficiency against socioeconomic status, demography, patterns of healthcare consumption, and other data. Perhaps we would discover determinants of efficiency, that we could change by some kind interventions and reduce inefficiency. A third strategy would be to look at the input side. We saw that the major cause of output shortfall was inefficiency, but a proportion was due to differences in inputs. If we cannot eliminate the inefficiency, perhaps we could offer more support to inefficient individuals as a compensation.

The third strategy is bound to involve the allocation of healthcare resources. This leads us on to prioritizing between different disease areas, since resources are limited, and we wish to spend them so that we maximize the benefits from their use. In order to prioritize between different disease areas, a common measure is needed, preferably scalar and based on patients' preferences [15]. One example is the QALY. An important limitation of our measure is that it is disease-specific, and specific to type 1 diabetes at that. Another way of looking at this could be to consider relative needs within a disease area. This is what our measure does. It does not help us judge the level of need between disease areas, but it does tell us what a person manages to accomplish compared to others in the same disease area (while adjusting for how

much inputs they use). I think this is person-centred in a very appealing way. Although it does not satisfy the need to be able to prioritize between different disease areas, it does provide another kind of information which is also important, and which could perhaps be used to carry out diabetes care. Perhaps this is a backward move from a scientific point of view, but it might be a direction still worthy of some attention.

4.3 Predicting future costs and risk factors

Among our PROM and PREM scales, we found only one predictor of future diabetes care costs. In type 1 diabetes, high score of *General wellbeing* predicted lower future inpatient care costs, stemming mainly from fewer inpatient admissions but also to some extent from shorter length of stay (data not shown). We found no predictors in type 2 diabetes. We had a rich data material comprising healthcare resources and costs, diagnoses, health data, PROM, PREM, demography, risk factors for diabetes complications, and socioeconomic status, all deterministically linked in the individual patient level. Our sample was not very large, 884 individuals in total. This may have prevented us from detecting relationships with costs, often associated with high variability. On the other hand, the one finding we made concerned inpatient care costs in type 1 diabetes, the category with the highest variability in our material (see Table 3 in Paper I).

Even though we made almost no findings of PREM and PROM predicting future costs, fair regression models for future costs could be fitted to our data. Their coefficients of determination varied, but some of the models ought to be useful for predicting costs, possibly after refinement. This was however not in the scope of study IV, but we could pursue this in the future (See Future Perspectives).

We did, however, find predictors of future HbA1c. We found that *Manage diabetes* predicted low future HbA1c in type 1 diabetes. This appears plausible. We also found that *Satisfaction with treatment* predicted lower future HbA1c in type 2 diabetes. Husdal et al found a slightly similar result, that primary care centres that used individualized treatment plans for their type 2 diabetes patients were associated with lower HbA1c levels [24]. Husdal et al asked the staff at the primary care centres how they organized their care, whereas our scale *Satisfaction with treatment* was based on individual patient data. But the respective scopes of their questions and ours ought to overlap at least a little since they address similar constructs. Husdal et al write that their study design prevents them from assessing causality. Our design with *Satisfaction with treatment* observed before HbA1c, and controlling for past HbA1c levels, lends at least some support for a causal hypothesis. We do however not go further than to say that *Satisfaction with treatment* predicts future HbA1c.

Thus, we found predictors of future HbA1c in both type 1 and type 2 diabetes. Lower future HbA1c ought to result in lower costs for managing complications, even though we did not detect this in the relatively short time frame of our study.

Why did we see predictors of HbA1c among PROM and PREM above, but no predictors of SBP and LDL? Already in studies I and II, we saw that there were few correlations with the risk factors among the PROM and PREM, those that we saw were mainly with HbA1c. Furthermore, we have argued that HbA1c can give symptoms, and an individual with diabetes has HbA1c in focus, but this is not the case with SBP and LDL [10]. Last, when the questionnaire in study II was developed, based on clinical experience, no strong correlations between scales and SBP and LDL were expected. The results also matched these expectations [11]. Also, in interviews serving to identify aspects important to individuals with diabetes, blood pressure and lipids never came up neither in connection with wellbeing nor with living with diabetes [19, 25].

When we examined diabetes care costs, it was unavoidable to think about our use of PREM as inputs in study III, and the opportunity in study IV to validate the PREM against resource consumption and costs. We did, in a precursor to the resulting study IV, to separate costs into production costs - managing diabetes and controlling risk factors - and consequence costs - managing diabetes complications. Other studies seem to have succeeded in making this separation [4, 5], as mentioned in the introduction. Unfortunately, we had to call our attempt a failure, since we saw inadmissible relationships such as support being negatively associated with production costs. Perhaps this was due to our PREM reflecting other things than reimbursed quantities, that our PREM corresponded to how the patient-staff dialogue was, rather than the length of the dialogue or the number of test or procedures etc carried out (that would have been reimbursed). Acquiring the ability to carry out a good dialogue, appreciated by the patient, and resulting in a favourable PREM rating, surely had a cost. But this cost may have been a one-time cost, never reimbursed, or at least not reimbursed per future visit. Another possible explanation is that the lack of admissible relationships was due to our cost data being based on flat-rate reimbursement models.

To summarize, we discovered predictors of HbA1c, the most important risk factor, among our PROM and PREM. We had hoped to discover predictors of future diabetes care costs too, so we were only partially successful. But we took a small step forward. If neither PROM nor PREM does predict future diabetes care costs, this is important knowledge, but more research is needed before such a conclusion can be made (if this conclusion can be drawn at all). In the introduction, I asked whether PROM and PREM could be used to evaluate diabetes care. This question may be answered positively by study III. In study IV the answer is partial; with regards to HbA1c but not yet with regards to costs.

4.4 The individual perspective

In this subsection, I list some observations from previous subsections that describe how I believe we contributed to the individual perspective, organized in a couple of categories. I also bring up one last category where we risk making a negative contribution to the individual perspective.

4.4.1 PROM and PREM: fundamental for the individual perspective.

In study I, we assembled a set of analytic procedures for developing IRT scales and determining the measurement properties of the scales. One contribution to the individual perspective, for which obtaining PROM and PREM must be considered fundamental, was carrying out the procedures and thus developing the PROM and PREM scales. The development of the second questionnaire, based on what individuals with diabetes themselves thought important and using their own words, was an essential contribution mainly belonging to Dr Svedbo's thesis [11], though the present thesis can claim part of the contribution, namely developing the PROM and PREM scales from this questionnaire. In addition, we learned of positive clinical experience of using these scales in the staff-patient meetings at test clinics. This is currently being studied, but preliminary results indicate that patients appreciate discussing their PROM and PREM with the staff. Here it should be noted that our PREM were carefully developed to address the individual patient's needs of support, access etc in very general terms to avoid focus on the staff, judgement of the staff or anything similar. I am quite proud that our collective work may have had something to do with this positive experience.

4.4.2 Quality of life using individual weights

We developed a measure of the quality of life that allowed individuals to use their own weighting of the involved variables. This design did influence the measure, which we saw in the heterogeneity in which variables contributed to the individuals' quality of life estimates. I would argue that this is an important contribution to the individual perspective, though some challenging methodology was involved, making the contribution perhaps a bit inaccessible. Further development, validation and application could perhaps help convey the methodology to a wider audience.

4.4.3 Individual needs, improvement potential and vulnerabilities

The PROM and PREM scales could be used in both study I and II to obtain individual estimates of abilities, barriers, worries, and judgments of diabetes care. This is important, as we could act on such values with the aim of improving those who had low values. In study III, when we projected every individual onto the frontier, we could estimate their individual output shortfalls. Based on the frontier as a model, this represents their improvement potential and it is individually estimated. Furthermore, we saw - though presented only on the clinic level - that output shortfall was mainly due to inefficiency. We outlined strategies to address this, by resolving inefficiency or by using more support to compensate.

From study IV, we learned of predictors of future HbA1c. This could be used to identify individuals vulnerable to high risk of diabetes complications, and measures could be taken. We also saw that the presence of mental health problems and renal failure were predictors. Surely this is not a surprise for diabetes care, but still worth mentioning as it is important for the involved individuals. Looking more carefully at the presence of mental health problems and *Manage diabetes*, we discovered that *Manage diabetes* was significantly lower among those about to have mental health problems the year after the questionnaire. It could perhaps be used to predict upcoming problems (see Future Perspectives).

4.4.4 Some pieces of the puzzle of planning future diabetes care

As mentioned above, in study IV we discovered predictors of future HbA1c, and we also discovered predictors of one future cost category. Though we had hoped for more findings, these may be some first pieces of the puzzle of planning diabetes care, perhaps even on the individual level. Looking at the scales that predicted future HbA1c, their effect was sensitive to whether we controlled for past HbA1c levels or not. This is important for understanding how the scales work and it could perhaps help to interpret the scores of an individual patient, and this could contribute to the dialogue with the patient. For instance, did the patient rate *Manage diabetes* low because of poorly controlled HbA1c recently? We could discuss the patient's skills and knowledge and perhaps encourage the patient that the ability is not the problem. Instead, we could concentrate on other factors explaining the past HbA1c level.

4.4.5 Groups of individuals

We saw that age groups and diabetes type differed in terms of PROM and PREM (Figure 3.1), and that treatment groups differed in type 2 diabetes (Figure 3.2). Quality of life in type 1 diabetes differed between age groups, and we saw that the contributions to the quality of life from SBP, LDL and *Mood and energy* differed

by e.g. gender and age groups. Age and duration were predictors in cost models in study IV. Even though groups, of course, are not individuals, we get some general information by looking at the relevant group data, and perhaps this could serve as a starting point when we focus on the individual.

4.4.6 The integrity of the individuals

In study II, we learned of positive patient experience of using individual PROM and PREM results as basis for discussion in the staff-patient meetings. In study III, we used PREM at the individual level. These results have not yet been used in the dialogue between patient and healthcare staff, and perhaps never will. Nor have we shown individual results to anyone (apart from anonymously in Table 3.2). There is, from a principal point of view, a potential issue with PREM on the individual level that deserves special attention. PREM are judgments of experiences of healthcare. It may not be comfortable to meet the healthcare staff if they just saw your rating -- let's assume that you gave a poor rating, and this was not well received by the staff. It is possible to make a negative contribution to the individual perspective if we use PREM on the individual level in such a way that the data are revealed. The PROM could be used on the individual level, but perhaps some care is needed regarding PREM. I wish to point out that the PREM items in the NDR questionnaire were developed with this potential issue in mind, so they address individual patient needs for support, access, etc, and try to avoid focus on staff, ratings and the like. We nevertheless need to be aware of the potential issue.

4.5 Generalizability

4.5.1 Generalization to other diabetes populations

Studies I and II provided estimates of scale scores of various PROM and PREM, one set in study I and another set in study II. The methodology used to develop these scales was generic, so any issue with generalization relates to the PROM and PREM scores, and in the case of study II, the aspects judged as important and hence included in the second questionnaire. The study samples in study I and II were judged representative of the NDR, and the NDR has a very high coverage rate, an estimated 94% [1]. Hence, every individual with diabetes in Sweden is probably well represented by our data, that is, if it weren't for our response rate of 60% in study I and around 55% in study II, among the initially invited individuals. Clearly, our results apply to respondents - this ought to be the case for any questionnaire survey. We need to be careful not to generalize our results to those better represented

by the non-responders. Here, for instance, we would probably find those having difficulties with a questionnaire in Swedish.

Having abandoned the questionnaire in study I in favour of the new questionnaire in study II might suggest that generalization of study I is uninteresting. But we have ideas in Future Perspectives that may speak of the opposite. Furthermore, the second questionnaire, being based on interviews of Swedish individuals, covers a set of aspects identified as important. Would these aspects be the same, had the study been performed abroad? This is difficult to know, but generalizability might be limited by different organization diabetes care, different financing of the healthcare system, and perhaps cultural differences.

Study III used a subset of individuals with type 1 diabetes from the study population in study II, selected by having registry data required for study III. This population is slightly smaller in size, and the difference in selection could in principle add further limitations to generalizability - from a mere population perspective. But other aspects need to be considered: a new population will give rise to another front. Adding people to the population will make the front either stay the same or expand (the front can never shrink with the old observations still there). With a new front, our current patients would see new effectiveness estimates, i.e. their quality of life would change. We could also keep the current front, but then we would probably see some observations with effectiveness > 1. This would happen if a new observation is made outside the current front. Observations outside the current front could be from individuals being more successful than we have ever seen before, or the "technology" could have changed to the better. Technology represents what can be accomplished at best, and a change could be e.g. a new more efficient drug. These are aspects of generalizability that originates from efficiency analysis. We could handle this with "version control" of the fronts, e.g. the estimates in study III relates to the year 2015 front (our data being from 2015). Next time, we either use the same front and accept seeing the occasional quality of life estimates > 1, or we use a new front. With the new front, comparisons to old results are no longer possible unless you weigh in both fronts. A Malmquist index could be used, it adjusts for both changes in efficiency and change in technology [16]. Perhaps one could think of different versions of fronts as a sort of parallel to (but not quite the same as) different value sets for health-related quality of life data.

Our study IV was based on a rather small sample, and it gave rise to few findings, so rather than generalizing the results, I would suggest carrying out a new larger study, perhaps with design changes. Apart from that, however, the study IV population was from the Västra Götaland region, and the costs were reimbursements from that same region. The reimbursement may not be the same in other regions of Sweden, so generalization may be somewhat difficult. On the other hand, the population of Västra Götaland region is not a negligible fraction of the Swedish

population and therefore ought to be fairly representative for Sweden, so one should not rule out generalizability entirely. The generalization to other countries would probably be difficult, since it is notoriously difficult to transfer economic studies between countries, stemming from differencs in e.g. demography and epidemiology, clinical practice, relative prices of resources and ways of financing healthcare [15]. The results regarding HbA1c would probably be easier to generalize across borders.

4.5.2 Generalizability to other disease areas

Development of PROM and PREM questionnaires is a wide field, and development of IRT scales is too. A considerable number of questionnaires have been developed in many disease areas. Our studies I and II used generic methods, and generalization is not so interesting, as mentioned above. Nor would generalization of study IV be, as mentioned above. Generalization of study III, on the other hand, might be interesting. Some of the PROM and PREM scales were diabetes-type specific, which prevented us from carrying it out for both diabetes types together. Therefore, type 1 was selected for a first study but an obvious extension would be to type 2 diabetes. Type 2 satisfy the prerequisites just as type 1 did. However, diabetes type-specific scales would prevent direct comparisons between type 1 and type 2.

Our measure of the quality of life in study III is based on how individuals generate outputs, relative to the most successful individuals (with the same mix of outputs). The measure being relative makes it difficult to extend such that it would enable comparisons across disease areas. Though this does not prevent generalization of the method itself to other disease areas. Maybe *application* is a better word than generalization, though.

From a principal point of view, I see similarities between diabetes and a couple of other disease areas. But the use of efficiency analysis to measure the quality of life does not appear to be widespread. I think that the approach can be applied more widely, for chronic conditions where there are treatable biomarkers that are risk factors for future morbidity, where these risk factors play a central role, and where the collection of patient-reported measurements would be feasible. With regards to the mechanisms at work in my application to diabetes, I see two parallels in other diseases. The first parallel is where a biomarker acts in a somewhat similar role as HbA1c, and the second parallel is where we have a trade-off between short-term wellbeing and future health.

In the management of Chronic Obstructive Pulmonary Disease (COPD), lung function is used to determine disease severity and is routinely monitored along with symptoms. The main goals of managing COPD is to reduce the deterioration of lung function and to reduce symptoms [26]. Furthermore, the choice of drug treatment is informed by the lung function level [27]. A major component of COPD care is self-

management, where the patient is encouraged to adopt healthy behaviour and to develop skills to manage their disease (abilities making an additional parallel with diabetes). In cancer, Myeloma and Prostate cancer serve as examples. Myeloma cannot currently be cured, but a patient can survive for a long time with recurrent periods of treatment as needed, and the disease course becomes similar to that of a chronic condition [28]. The myeloma protein *M component* is a risk factor for progression [28]. In Prostate cancer, the prostate-specific antigen (PSA) level and how it changes over time is used to inform treatment decisions [29]. PSA is also monitored after completed treatment to predict recurrence and metastatic disease. Thus, we have treatable risk factors that indicate future health. An application of the quality of life measure would make use of PROM to measure wellbeing, and lung function, the M component, or the PSA level, respectively, to measure future health.

I see parallels of the second kind in cancer. A first example is in breast cancer. If the cancer responds to hormonal therapy, then hormone blocking medications could help to slow down or even stop the breast cancer cell growth [30]. However, some patients experience side effects, making them choose between (a) skipping treatment, giving better wellbeing for the moment but some risk of relapse of breast cancer, or (b) taking the treatment, and risk having poorer wellbeing but having a reduced risk of relapse. A second example is when individuals have genetic factors that (almost surely) cause cancer. For instance, a lack of DNA mismatch repair genes, associated with a high risk of colorectal cancer. The individuals might choose to ignore this and risk getting cancer. Or they could accept being followed-up biannually and hereby likely prevent cancer. Though this needs to go on, year after year, with cumbersome examinations, and the individuals are constantly reminded of their risk. An application of the quality of life measure in these two examples would use PROM to measure wellbeing and let the individual weighting play an important role. Perhaps a positive experience of the follow-up procedures can lessen the discomfort of going through the examination, thus PREM could play an interesting role here too. The biomarkers may not play such a prominent role here as in diabetes. Though perhaps some imaging biomarker indicating disease progress could be used.

To apply the measure in a new disease area is one thing. Perhaps we could also transfer knowledge in other ways. We saw in study IV that the presence of other illnesses affected future costs. This is important knowledge for those managing patients with multiple morbidities. In addition, it should be communicated to diabetes care as well as e.g. cancer care and psychiatric care. E.g. by encouraging networks involving e.g. diabetes nurses, cancer contact nurses and so one.

4.6 Policy implications

Any potential policy implications from the work in this thesis appear to fall into three categories.

First, can we discover individuals' need, and can diabetes care react accordingly?

- Perhaps the most promising way to use this is to use the PROM and PREM scores in the dialogue with the patient. The clinical experience is good, as well as patient experience, however one need to keep in mind the potential concerns regarding PREM on the individual level (see section 4.4.7).
- We discovered predictors of future HbA1c, and perhaps this can be used to detect individuals vulnerable to an elevated risk of complications in the future. We could use this to trigger interventions. We do not need to use data on previous costs as predictors, which I think may be a relief for diabetes care.

Secondly, we could apply our results to evaluate alternatives and allocate resources.

- The PROM and PREM scales can be used to measure and characterize the alternatives, and this can inform choices between them.
- We can measure the quality of life, which could further inform choices between alternatives. Using it for resource allocation could help to maximize the production of quality of life e.g. under resource constraints.
- Being able to measure the quality of life is a fundamental basis for understanding reasons for the poor quality of life. Knowing the causes would be a first step towards intervening on such causes.
- Our measure of the quality of life weighed in the available inputs. An alternative to intervening on causes of inefficiency is to allocate more inputs to compensate for poor efficiency that leads to poor quality of life.
- The quality of life measure is however in its current application limited to use in type 1 diabetes, and certainly, another limitation is that the measure being based on diabetes-related PROM and PREM scales cannot be used outside the field of diabetes.

A third category is learning from best practice.

- Being able to measure the quality of life and looking at the inputs and outputs that generate quality of life, we ought to be able to discover not only which diabetes caregivers are successful but also what they do to succeed. Perhaps we can use this to transfer good practice to other caregivers. Our scales and quality of life measure could then help us evaluate if this has a good effect.

5. Conclusions

We have gathered a set of procedures for assessing measurement properties of PROM and PREM scales, and we accomplished collection of PROM and PREM from individuals with diabetes.

The measurement properties of the PROM and PREM improved further with a new more patient-centred questionnaire. The PROM scales describe abilities and freedom from worries and barriers, and the PREM scales describe how individuals experience various aspects of support from diabetes care.

We could use the scales to describe our study populations, and compare groups, and identify vulnerable individuals.

PROM, PREM and risk factors of diabetes complications were importantly found to be complements to each other in describing the situation of an individual with diabetes, and all three types of data were found to be useful.

We could measure the quality of life, and the measure allowed every individual to use their own importance weights for the variables involved. The method also allowed us to estimate the individual's improvement potential.

One PROM scale and one PREM scale were found to predict future HbA1c levels, and their effect is sensitive to past levels of HbA1c.

We have contributed to the individual perspective in diabetes in several ways, and we have identified a couple of different approaches to develop diabetes care, e.g. through the study of causes of poor quality of life.

The PROM and PREM are now used in day to day clinical practice in Sweden.

6. Future perspectives

In this section I list several items that I find worthy of consideration for future research, organized in three categories: (a) follow-up of scales, (b) further work on the quality of life measure, and (c) predicting future outcomes.

6.1 Follow-up studies

When publishing the study in 2014, we estimated that some accumulated 20 000 patient-years of observation time had passed since we obtained the PROM and PREM data in study I. To date, this would have increased substantially. This would make a formidable case for a register-based follow-up of survival, or diabetes complications -- data are readily available for such research, to study whether any of the PROM or PREM scales are predictors of such events.

Regarding study II, its PROM and PREM data was collected in 2015, so only five years has passed since. Still, some 18 000 patient-years of observation may have accumulated, so here would be another case of register-based follow-up, should the scales in the second questionnaire be predictors of survival or complications.

6.2 The quality of life measure

There are a couple of research questions that would appear almost unavoidable. Extending the quality of life measure to type 2 diabetes is one. I find it likely that a straightforward application of the current quality of life production model would work equally well, though separate scales for type 1 and type 2 diabetes would result in separate quality of life measures. How to relate to these separate measures - or indeed figure out how to make comparisons - is another tempting case of future research.

I have mentioned application in cancer above, that would also make interesting extension of the work, not only including parallels in another disease area, but applications where the risk factors have somewhat different roles. Again, interesting questions of comparability would arise.

The role of PREM in the quality of life production model could be discussed, whether the PREM scales might be outputs as well in the capability model.

Preliminary findings (deliberately not shown), inspired by findings in study IV, indicate that this might not only work but would even improve the model.

Another interesting prospect for research is a phenotype model: We though that some scales might play different roles for different individuals: *Freedom from worries* might be an ability resembling resilience against worries, or it could be a favourable outcome. Would it matter how we used *Freedom from worries* in the production model of quality of life? I believe it would. A phenotype model is a model that allows for different parameter sets for different subgroups of individuals, e.g. whether a scale is an input or an output. The model is self-organizing, using statistical methods to determine which phenotype an individual is likely to belong to, and estimating the parameters for a phenotype based on the likely member individuals. This is iterated until convergence is obtained, and the procedure identifies the number of different phenotypes and the parameters that define each phenotype [31].

It would be unavoidable to mention validation of the quality of life measure. We see several options. We could conduct a time trade-off or a standard gamble study, or perhaps even a willingness-to-pay study, to validate our measure. Or we could use an established instrument such as the EQ-5D (we did use EQ-5D, but with the old questionnaire, not with the new questionnaire upon which we base our quality of life measure).

6.3 Future outcomes

We saw in study IV that the cost models achieved quite some coefficient of determination in some cases. These models could perhaps be further refined, to be useful for predicting future costs. This appears useful for e.g. diabetes care planning, or modelling and evaluating new interventions. Though there are no shortage of health economic models for evaluating interventions in diabetes, e.g. there are the IMS Core model and the IHE models. Perhaps another Swedish model would be appealing at least locally, as a standalone model or perhaps as a new cost module in an existing model.

In study II, diabetes caregivers were not aware of the scale scores. One might design a study where they would be made aware of scale scores. Them knowing the scores and acting on them could have an impact on future outcomes, and this could be evaluated. This might indicate how useful the scales are for diabetes care, or how much use diabetes care can make of the scales.

We made the preliminary finding that the PROM scale *Manage diabetes* was a potential predictor of future mental health problems. This could be valuable information for diabetes care (and psychiatric care) but needs to be properly studied.

7. Acknowledgements

This thesis has a single author, but it really is the result of teamwork. I pride myself with having the ability to surround myself with great people, and I have often found that I work in excellent teams. Perhaps it is luck rather than an ability. Regardless, a lot of people have supported my work with this thesis. First of all my family: Klara and Jenny (-- when you went to school, I did too!) and most of all Annette: always encouraging and after half a decade or so, perhaps a bit impatient. My supervisors Ulf Gerdtham and Soffia Gudbjörnsdottir have contributed enormously to my research education, but also in wider perspectives. I appreciated having Stefan Lindgren and Tobias Rydén as supervisors the first few years, before I narrowed the thesis down to the Diabetes field. Anna Glenngård has had very important impact in supporting, reviewing, suggesting, and not least in being brave enough to point out stupid ideas. Bo Palaszewski has provided scientific expertise and many interesting discussions on a wide range of subjects, in a way placing a plimsoll mark on my research. Maria Svedbo for discussions of life, research and survival, and not least, our respective studies. Katarina Eeg-Olofsson for providing vital bits of the diabetes puzzle, and supervising my research in general, despite not formally being a supervisor. Ulf Persson put up with me for a long time and encouraged me to start with the thesis. Dorry Wunderlich contributed with rapid thinking, and a great sense of humour. Gunnel Hedberg provided many interesting perspectives and made quite a cultural imprint. Björn Larsson introduced me to "björnpop" (music ideal for when you try to concentrate on challenging scientific work) and we have had many interesting discussions of science, art and more. So many other friends supported simply by being there, providing links back to reality. It has been a joy to have all the colleagues at the Health Economics Unit at the university, at the Regional Cancer Centre South (thank you again for the combined harvester toy!), at the National Diabetes Register, at the Swedish Institute for Health Economics (source of important literature on *virtual professions* - thank you Annette!), and elsewhere, giving professional support and interesting discussions. Anna Lindqvist came to my rescue when difficult numbers (0-9 in various troubling combinations) rebelled against me. Ulrika Larsson helped me develop strategies to understand, manage and cross some particularly challenging barriers. Thank you Gunilla Albertén for turning this thesis into a proper book. The helpful staff at the medical faculty should be mentioned as well. I would like to acknowledge the Swedish Institute for Health Economics (IHE) for providing the platform for undertaking doctoral studies to me

as an employee (the first years), the Swedish National Diabetes Register for being my main sponsor, provider of data and expertise, and most of all for bestowing upon me the quest to carry out the studies in this thesis, and also Abbott (Thomas Lundquist) for sponsoring activities early in my doctoral studies that unfortunately weren't included in the thesis. Last, but not least, all of you who filled out the NDR questionnaires and provided me with data for my work. Thank you. I hope our work will provide value to you and others who live with diabetes.

References

1. Nationella Diabetesregistret, *Årsrapport 2018*. 2019, Nationella Diabetesregistret: Gothenburg.
2. Häger Glenngård, A. and S. Borg, *Can people afford to pay for health care? New evidence on financial protection in Sweden*. 2019, WHO Regional Office for Europe: Copenhagen.
3. Bolin, K., et al., *Diabetes, healthcare cost and loss of productivity in Sweden 1987 and 2005--a register-based approach.* Diabetic medicine : a journal of the British Diabetic Association, 2009. **26**(9): p. 928-34.
4. Jonsson, B., *Diabetes--the cost of illness and the cost of control. An estimate for Sweden 1978.* Acta Med Scand Suppl, 1983. **671**: p. 19-27.
5. Henriksson, F. and B. Jonsson, *Diabetes: the cost of illness in Sweden.* J Intern Med, 1998. **244**(6): p. 461-8.
6. Norlund, A., et al., *Cost of illness of adult diabetes mellitus underestimated if comorbidity is not considered.* J Intern Med, 2001. **250**(1): p. 57-65.
7. Borg, S., et al., *Patient-reported outcome measures and risk factors in a quality registry: a basis for more patient-centered diabetes care in Sweden.* Int J Environ Res Public Health, 2014. **11**(12): p. 12223-46.
8. Borg, S., et al., *Patient-reported outcome and experience measures for diabetes: development of scale models, differences between patient groups and relationships with cardiovascular and diabetes complication risk factors, in a combined registry and survey study in Sweden.* BMJ Open, 2019. **9**(1): p. e025033.
9. Borg, S., et al., *Quality of life in chronic conditions using patient-reported measures and biomarkers: a DEA analysis in type 1 diabetes.* Health Econ Rev, 2019. **9**(1): p. 31.
10. Borg, S., et al., *Patient-reported measures and the relationships with costs and risk factors for complications in diabetes.* In preparation.
11. Svedbo Engstrom, M., *Patient Perspectives brought to the Fore for Diabetes Care. Descriptions as well as Development and Testing of the Diabetes Questionnaire*, in *Department of Molecular and Clinical Medicine Institute of Medicine*. 2019, Sahlgrenska Academy, University of Gothenburg: Gothenburg, Sweden.
12. Burstrom, K., et al., *Swedish experience-based value sets for EQ-5D health states.* Qual Life Res, 2014. **23**(2): p. 431-42.
13. Nyqvist-Persson, M., *Personal communication*. 2019: Regional Cancer Centre South, Lund, Sweden

14. EuroQol Group, *EuroQol--a new facility for the measurement of health-related quality of life.* Health Policy, 1990. **16**(3): p. 199-208.
15. Drummond, M.F., et al., *Methods for the economic evaluation of health care programmes*. 2nd ed. Oxford medical publications 1997, Oxford ; New York: Oxford University Press. 305 p.
16. Bogetoft, P. and L. Otto, *Benchmarking with DEA, SFA, and R*. International series in operations research & management science,. 2011, New York: Springer. xvi, 351 p.
17. Cherchye, L., et al., *An introduction to 'benefit of the doubt' composite indicators.* Social Indicators Research, 2007. **82**(1): p. 111-145.
18. Färe, R., et al., *Evaluating health care efficiency*, in *Advances in Health Economics and Health Services Research*, J.L.T. Blank and V.G. Valdmanis, Editors. 2008, Elsevier: Amsterdam. p. 209-28.
19. Svedbo Engstrom, M., et al., *What is important for you? A qualitative interview study of living with diabetes and experiences of diabetes care to establish a basis for a tailored Patient-Reported Outcome Measure for the Swedish National Diabetes Register.* BMJ Open, 2016. **6**(3): p. e010249.
20. Zethelius, B., et al., *A new model for 5-year risk of cardiovascular disease in type 2 diabetes, from the Swedish National Diabetes Register (NDR).* Diabetes research and clinical practice, 2011. **93**(2): p. 276-84.
21. Cederholm, J., et al., *A new model for 5-year risk of cardiovascular disease in Type 1 diabetes; from the Swedish National Diabetes Register (NDR).* Diabetic medicine : a journal of the British Diabetic Association, 2011. **28**(10): p. 1213-20.
22. Rawshani, A., *Epidemiological aspects of cardiovascular morbidity and mortality among individuals with diabetes: the relative importance of cardiovascular risk factors*, in *Institute of Medicine. Department of Molecular and Clinical Medicine*. 2018, University of Gothenburg. Sahlgrenska Academy: Gothenburg.
23. Mansdotter, A., et al., *We Propose a Novel Measure for Social Welfare and Public Health: Capability-Adjusted Life-Years, CALYs.* Appl Health Econ Health Policy, 2017. **15**(4): p. 437-440.
24. Husdal, R., et al., *Associations between quality of work features in primary health care and glycaemic control in people with Type 2 diabetes mellitus: A nationwide survey.* Prim Care Diabetes, 2018.
25. Svedbo Engstrom, M., *Personal communication*. 2020: School of Education, Health and Social Studies, Dalarna University, Falun, Sweden.
26. Global Initiative for Chronic Obstructive Lung Disease, *Global Strategy for the Diagnosis, Management and Prevention of chronic obstructive pulmonary disease*. 2018, Report. On line. URL=http://www.goldcopd.org (Accessed on April 20, 2018).
27. Läkemedelsindustriföreningen, *Farmaceutiska Specialiteter i Sverige (FASS). On line, URL=www.fass.se, Accessed 26 October 2018.* 2018.
28. Regionala cancercentrum i samverkan, *Myelom. Nationellt vårdprogram*. 2018, On line. URL=www.cancercentrum.se, Accessed 26 October 2018.

29. Regionala cancercentrum i samverkan, *Prostatacancer. Nationellt vårdprogram*. 2017, On line. URL=www.cancercentrum.se, Accessed 26 October 2018.
30. Breastcancer.org. *Hormone Receptor Status*. 2020; Available from: www.breastcancer.org.
31. Borg, S., et al., *Cost-Effectiveness and Heterogeneity: Using Finite Mixtures of Disease Activity Models to Identify and Analyze Phenotypes*. 2015, Working Paper 2015:5. Department of Economics. School of Economics and Management, Lund University: Lund, Sweden.