Patient Reported Outcome Measures

in Interventional Cardiology:

Listen To The Patient

Mr. Ian Kemp

A thesis submitted in partial fulfilment of the requirements of

Liverpool John Moores University for the degree of

Doctor of Philosophy

March 2021

CONTENTS

Table of Figures

List of Tables

Abbreviations

AF	Angina Frequency
AS	Angina Stability
CABG	Coronary Artery Bypass Graft
CAD	Coronary Artery Disease
CCS	Canadian Cardiovascular Society
CHD	Coronary Heart Disease
CHP	Cardiac Health Profile
CI	Confidence Interval
CVD	Cardiovascular Disease
DES	Drug Eluting Stent
DoH	Department of Health
ED	Erectile Dysfunction
EQ-5D	EuroQol Group
GP	General Practitioner
HDAS	Healthcare Database Advanced Search
IHD	Ischaemic Heart Disease
IQR	Interquartile Range
LVF	Left Ventricular Function
MACE	Major Adverse Cardiac Events
MI	Myocardial Infarction
MOS	Medical Outcomes Study
NHS	National Health Service
OMT	Optimal Medical Therapy
ONS	Office of National Statistics
PCI	Percutaneous Coronary Intervention
PGA	Patient Global Assessment
PhGA	Physician Global Assessment
PL	Physical Limitation
PRO	Patient Reported Outcome
PROM	Patient Reported Outcome Measure
QALY	Quality Adjusted Life Year
QL	Quality of Life
QoL	Quality of Life
SAQ	Seattle Angina Questionnaire
SoS	Stent or Surgery (trial)
SPSS	Statistical Package for the Social Sciences
TS	Treatment Satisfaction
UK	United Kingdom
US	United States
USA	United States of America
VAS	Visual Analogue Scale
WHO	World Health Organisation

Abstract

Patient Reported Outcome Measures (PROMs) are instruments that seek patients' views on symptoms, functional status and health related quality of life. My thesis analyses data from the Stent or Surgery Trial. A cohort of 988 patients completed several PROMs before, and at two time points after their revascularisation procedure, and this data forms the basis of my thesis.

I found discordance between patients' and their physicians with physicians slightly overstating angina burden prior to revascularisation and under reporting at twelve-months post procedure by comparison to patients.

I compared the generic EuroQol EQ-5D and disease specific Seattle Angina Questionnaire (SAQ) to assess the degree of agreement. I found the EQ-5D visual analogue scale (VAS) had poor correlation with any of the five domains of the SAQ. However, the VAS relates to how a patient feels at the time of completion and the SAQ refers to the previous four weeks.

I then compared the full 19 question SAQ with a short 7 question version (SAQ-7). A single summary score was calculated for both instruments. Each timepoint showed a strong positive linear correlation above 0.9. The implication of this finding is there may be greater take-up for a PROM with fewer questions and that a single summary score is easier to understand.

Finally, I examined some strengths and weaknesses of questionnaires. Advantages include expense, practicality, speed and anonymity. Disadvantages included user fatigue, dishonesty and bias in the way the questionnaire is designed.

PROMs in the UK tend to be used to compare performance between individual operators or Trusts. In other countries such as Sweden they are used to calibrate individual patients' treatment plans.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Acknowledgements

I would like to express my gratitude to my Clinical Supervisor, Professor Rodney Stables and my Director of Studies at Liverpool John Moores University, Professor Paulo Lisboa. Without their patience, wise guidance and encouragement this thesis would not have been completed.

I am also grateful for statistical guidance from Dr Steve Lane at Liverpool University.

I would also like to acknowledge my late parents, Roy and Elizabeth for a lifetime of support, love and guidance. How I wish they were still here to see me now.

Chapter 1 Introduction

"*The ultimate measure by which to judge the quality of a medical effort is whether it helps patients (and their families) as they see it. Anything done in healthcare that does not help a patient or family is, by definition, waste, whether or not the professionals and their associations traditionally hallow it*" [(1)]

For many years the impact of healthcare interventions was measured in terms of the ability to limit mortality [(2)]. Mortality data is provided in detail by the Office of National Statistics (ONS) via its website at www.ons.gov.uk [(3)]. A statistical bulletin in 2013 states that infant mortality decreased, and that cancer was the most common broad cause of death accounting for 29% of all deaths registered. The bulletin states the number of deaths recorded is related to the size of the population and its age distribution. We can see that 28% of all deaths were due to circulatory diseases such as strokes and heart disease. The bulletin also states that over the last century there have been steady decreases in mortality rates of three main causes of death (cancer, circulatory and respiratory). The bulletin went on to say this is partly due to improvements in treatment and diagnosis.

Several other factors contributed to reductions in mortality. Cutler and Meara [(4)] make the point that all-cause mortality was in decline well before modern medical treatment was available. The reductions could be related to nutritional improvements and public health measures. In the first part of the 20th century there were advances in the treatment of infectious diseases and the introduction of penicillin. Cutler and Meara [(4)]state "*The cumulative decline in deaths from infectious diseases from 1900 to 1960 was 92%*". They also state the cumulative decline in cardiovascular disease mortality between 1960 and 1995 was close to

two-thirds. This reduction in cardiovascular mortality could be attributed to several factors, including drugs to dissolve blood clots, surgical procedures such as bypass surgery, angioplasty, coronary care units, trained response teams and pharmaceuticals. Other factors that may contribute to reductions in mortality include changes in behaviour such as reduced rates of smoking, socioeconomic factors such as better education or increased aspirations for better health [(4)]

Education campaigns have influenced people's diet, lifestyle and smoking habits, an example being a campaign called '*Stoptober*' which encouraged people to stop smoking for the month of October[(5)]. However, Devlin and Appleby [(2)] point out that only a small proportion of people being treated in health systems in any given period actually die. Most receive some improvement due to their treatment. Devlin and Appleby point out that publishing data on reinfections, reoperations and adverse events tell us nothing about the patient *experience* of healthcare. One of the most important, yet difficult, questions in clinical research is to identify the outcome measures that might best provide insight into the impact of the intervention being tested.

Mortality data is, nevertheless, useful for several reasons. These include population estimates, examination of suicide rates or drug related deaths, analysis of infant mortality and comparison against the census estimates each decade. These epidemiological analyses inform estimates of trends at a population level. However, they tell us nothing about the patient *experience* of healthcare.

The evaluation of services tended to focus on efficiency of service delivery or the range of services on offer [6]. These measures - although important - say nothing about the patient's own health. Lord Darzi's interim report on the future of the National Health Service (NHS) in 2007 recommended that patient reported outcomes should have a greater role [7]. In the NHS Contract for Acute Services in April 2008, Patient Reported Outcome Measures (PROMs) were recommended for hip and knee replacements, groin hernia and varicose vein surgery [8].

An example of the difference in clinical outcomes as opposed to the patient experience would be the treatment of prostate cancer. Penson in 2001 [9] stated that following nerve-sparing radical prostratectomy the rate of impotence varies between 32% to 70%. Clinically, following surgery there will be an increase in five-year survival but at the cost of loss of sexual function and erectile dysfunction (ED). Penson's report was a broad review of several health related QoL instruments used in ED and pointed out that function is generally assessed but 'bother' is not. Function was stated to reflect the degree of symptoms, whereas bother reflects the extent the symptoms impact the patient. My criticism of this area of research is patients can exaggerate or understate their usual sexual activity or failure to achieve erections. This may be because of embarrassment or shame or lack of understanding of the terminology. Other reasons for what Chang et al call 'factors affecting patient responses' could be the patient's background, personality, age or level of education [10].A second criticism is the low incidence of reports of the use of oral erectile aids such as sildenafil (Viagra), at 9%. Today that figure may be significantly higher, reducing the reliance on vacuum devices or penile injections.

Helgason et al in 1996 [11] specifically examined the distress for prostate cancer patients. They compared the disease population with a similar reference population. They were asking if the patients would trade-off long-term survival for intact sexual function. In all ages, 19% of 299 patients would not accept treatment and 28% would only accept treatment if it prolonged life by ten or more years. The conclusion was that waning sexual function was the most common disease-specific reason for distress. A key finding was treatment was known to prolong life expectancy but was *not necessarily what the patient would choose given the option*. My criticism is this study relied on an instrument called "*The Radiumhemmets Scale of Sexual Function*". The instrument was developed through successive in-depth interviews with just 30 prostate cancer patients. This group may not have been a representative sample of the population with prostate cancer.

In 2015-16 the United Kingdom (UK) spent £220 billion on health and social care including benefits to support people with disabilities, or 11.5% of UK national income or 28.7% of total public spending[12]. With this level of spending on health care, the question arises, what is *produced* and is there *value* for money? Other areas of the economy can have their output measured, such as industry like steel works. Costs are evaluated and profits generated. However, health services pose a challenge in terms of measuring '*output*'. It is possible to count and cost the products used, such as equipment or drugs, or the cost of a stay in hospital. However, these are *intermediate* costs. Health services are not valued in their own right but by their *effects* on health[2]. The purpose of the NHS is not just to reduce harm but to promote health and social benefits in society as a whole.

1.1 Cardiovascular Disease Burden

Roth et al in 2017 reported data on cardiovascular disease (CVD) in the Global Burden of Diseases Injuries and Risk Factors 2015 Study [13]. They state CVDs are a leading cause of death in the world and a major barrier to sustainable human development. The number of CVD deaths per 100,000 of the population in Central Europe is 338, in high-income Asia Pacific it is 112 (the lowest) and in Central Asia it is 545 (the highest). Globally, the average is 286 per 100,000 of the population. Ischaemic heart disease (IHD) is a lack of blood to the heart muscle. IHD deaths per 100,000 of the population are 181 for Central Europe, 45 for high-income Asia Pacific (the lowest) and 336 for Central Asia (the highest). The global average is 142. The study reported that CVD deaths were declining in many high-income countries. However, the report also commented that the reduction in mortality from CVD has now reached a nadir and is no longer decreasing in high-income countries. The report authors speculate there may be several reasons for the flattening in the rate of decline in CVD deaths in developed countries, such as increases in obesity and air pollution and even changes in average air temperature although the authors state more research is needed in these areas. A criticism is this data is now several years out of date.

Bhatnagar et al reported in "*Trends in the epidemiology of cardiovascular disease in the UK*" that CVD was the second main cause of death [14]. There has been a 52% reduction in death rates between 1990 and 2013 and a 68% decline between 1980 and 2013. However, the authors note there has been little evidence of change in the prevalence of CVD in recent years. The report stated

that the number of patients registered at General Practice (GP) surgeries with coronary heart disease (CHD) has remained constant at around 3% of the population in England between 2004/05 and 2014/15. It also shows the number of hospital admissions for CHD in England has reduced from 428,262 in 2005/06 to 401,007 in 2013/14. If you include all diseases of the circulatory system, including strokes, then the number of admissions increased from 1,244,004 in 2005/06 to 1,401,232 in 2013/14. Bhatnagar et al[14] also reported that by 2013 there was a seven-fold increase in the number of percutaneous coronary interventions (PCI) than 1993. The number of coronary artery bypass grafts (CABG) peaked in the 1990s but has since decreased by 33% as PCI is now the dominant treatment option.

Luengo-Fernandez et al reported on the financial cost of CVD in the UK [15]. They state that CVD caused 40% of all deaths in the UK in 2012 (which includes strokes). The report authors attempted to estimate the economic cost of CVD and include healthcare costs, productivity losses and informal care costs. Informal case costs were defined as the opportunity cost of unpaid care. Included in healthcare costs were community health and social services, accident and emergency, day cases, inpatient stay, cardiac rehabilitation, outpatient services and drug costs. The estimated cost to the NHS in 2004 for CVD which includes strokes was £15.7 billion, or 21% of overall NHS expenditure. CHD was estimated to cost the NHS £3.45 billion in 2004, with 70% of this taken up by hospital inpatient care. The estimate for working days lost due to CVD incapacity was 69,346,572. The authors estimate the overall costs of CVD to the economy was £29.1 billion. They state the overall CVD related healthcare cost to the NHS

is 21% of the annual budget. It is the highest of any country in the European Union (EU). This cost was only surpassed by the estimated cost of mental illnesses. The authors stated the productivity loss costs were higher for CHD than for cerebrovascular accident (CVA) because more people die from CHD than CVA and that people die at a younger age from CHD than CVA. A criticism of the report is both the age of the data and the difficulty of estimating informal costs. The authors state themselves the monetary value of the time carers forgo to provide unpaid care is not available directly. Some of this cost can be inferred from the proportion giving care above and below the age of 65 years, roughly the age of retirement, but this may be wildly inaccurate.

1.2 Comparison between PCI and CABG

Gershlick and Thomas in 2007 stated that there have been major changes in the management of symptomatic obstructive coronary artery disease in the last 10 years [16]. They state there has been a shift towards PCI and there has been debate about which is "better" by clinical outcomes and overall cost estimates. Weintraub et al in 2008 [17] for the COURAGE trial randomly compared 2287 patients with stable coronary artery disease (CAD) to PCI plus optimal medical therapy (OMT) or to OMT alone. They showed that both groups showed improvements in health status. However, Gershlick and Thomas point out randomisation was performed *after* angiography and that most patients in the UK have angioplasty *because* of continuing symptoms despite OMT. They also pointed out that patients with left-main stem disease and left-ventricular dysfunction were excluded from the COURAGE trial and, therefore, if a similar

comparison was comparing CABG and OMT, there may be similar results. Another factor to be taken into account when comparing PCI and CABG is pre-procedural assessment, which tends to be more rigorous before CABG [16]. It was also noted that patency of grafts falls over time. Finally, Gershlick and Thomas point out several trial showing an 'advantage' for patients undergoing CABG were in the pre drug eluting stent (DES) period.

Prof. Surruys et al reported the five-year outcomes of the ARTS trial [18]. This trial randomised 1205 patients with multi-vessel CAD to PCI or CABG. This trial reported no significant differences in mortality, strokes or myocardial infarction (MI) between the two groups. However, there was a significant difference in major adverse cardiac and cardiovascular events (MACE), with 30.3% revascularisations in the PCI group and 8.8% in the CABG group. A criticism of this report is the age of the data, collected between 1997 and 1998, over twenty years ago. Since then, there have been developments in drug-eluting stents, catheter lab equipment and procedures, pharmacology and patient education. These factors will impact the differences in outcomes between the two allocated arms.

Cohen et al reported the final results of the SYNTAX trial [19]. This included patients with three-vessel or left main CAD. A total of 1800 patients were randomised to CABG or drug eluting stent (DES) DES-PCI. Over a five-year period, follow-up expenses – using American costs - were higher for the PCI group than the CABG group. This was caused by more frequent revascularisations and higher medicine costs. The conclusion was that CABG was both clinically and

economically advantageous. An obvious criticism is the use of American costs and the sole quality of life instrument was the EuroQol EQ-5D. However, this was used to calculate utility using a USA population. A full explanation of the EQ-5D is in chapter 3.

1.3 PROMs

Patient Reported Outcome Measures (PROMs) are defined by Black as instruments that seek to ascertain patients' views on their symptoms, functional status and health related quality of life [20]. Griggs et al [21] argue that PROMs "*…have the potential to be as valuable to the clinical encounter as the stethoscope is to the physical examination*". They argue that there is potential in PROMs to restructure the clinical encounter and enhance accuracy of prognosis and help identify at risk individuals. By "at risk individuals" they mean patients whose PROM data is significantly different from the norm for that patient group. This could mean patients who are depressed or experiencing greater pain than the average reported by that clinical group. They argue that although there is always the potential for using PROMs for regulatory and administrative purposes, the clinical potential should not be overlooked. They conclude that the dual goals of value-based and patient-centred care are achievable if PROMs are embraced by clinicians and patients. A criticism of the report is that it was vague about the practicalities of incorporating PROMs in day-to-day clinical practice. This would be a massive change in culture and require huge investment.

The United States has recommended the use of patient reported outcomes (PRO) in clinical trials and has stated "*the use of PRO instruments is part of a*

general movement toward the idea that the patient, properly queried, is the best source of information about how he or she feels" [22].

1.4 Patient-Physician Discordance

Patients' and their physicians' may have a different understanding of the patient's quality of life. This discordance can be elicited from the use of PROMs data and comparing this to the physicians' understanding and assumptions of the patients' quality of life. Douglas et al reported on patient-physician discordance in goals of patients with advanced cancer [23]. The study asked patients and their oncologists to complete a questionnaire at enrolment and every three-months until either the patients' death or fifteen months. A simple visual analogue scale (VAS) of 100 points was used, with zero the worst and 100 the best quality of life. At baseline, 24% discordance was observed, and for patients who survived at their last interview there was also 24% discordance. At the last interview before death for patients who did not survive to fifteen-months there was 28% discordance whereas at first interview had been 70%. The authors concluded there was a discrepancy between the goals of care of both oncologists and their patients. There was a deficiency of understanding between patients' and their oncologists and this reflected a lack of what they called "*high-quality communication and decision making*" [23]. The authors conclude that the use of a simple tool such as VAS could lead to enhancing the communication between oncologist and patient and this in turn could lead to more appropriate treatment decisions and incorporating the patients' and oncologist goals into the care plan. My criticism of this report is it was a single centre and enrolled just 11

oncologists. The problem with single centre studies if the generalisability of the results. If the 11 oncologists were all world leaders in their field this would not represent the average.

1.4.1 Measuring Discordance: Cohen's kappa

There are several methods for measuring the agreement between two raters, but the most common is Cohen's kappa [24]. Cohen's kappa measures the agreement between raters but takes into account the possibility of chance agreement. Kappa is intended to give the reader a quantitative measure of the magnitude of agreement between raters [25]. If a study measures a variable with only two possible states, such as dead or alive, then there will be high reliability between the raters. However, when there are finer discriminations reliability between the raters is harder to obtain. "*Reliability of data collection is a component of overall confidence in a research study's accuracy*" [24]. Kappa score can range from -1 to +1, where zero represents the agreement expected by chance, 1 represents perfect agreement .

The formula for kappa is: **k = Pr(a) – Pr(e) divided by 1- Pr(e)**

Pr(a) represents the actual observed agreement and Pr(e) represents chance agreement.

Weighted kappa can be used in cases where there are multiple categories. A weighted kappa can assign less weight to agreement as the categories are further apart. This means for example that a rating of zero by one rater and five by another could be assigned less weight than a rating of two by one rater and three by the other.

Cohen stated the kappa results could be interpreted as follows: ≤0 no agreement, 0.01 to 0.20 none to slight; 0.21 to 0.40 fair; 0.41 to 0.60 moderate; 0.61 to 0.80 substantial and finally 0.81 to 1.00 almost perfect. Kappa is affected by prevalence of the findings under consideration. With rare findings, very low kappa values may not necessarily reflect low rates of agreement [25].

Chapter 4 will discuss discordance in patient-physician reporting of the burden of angina experienced by patients.

1.5 Categories of PROMs

The systematic collection of PROMs data in the NHS began in 2009 and represented a recognition of the patient's perspective in both the quality and effectiveness in health care[2]. Initially data was collected in respect of four procedures, namely hernia repair, hip and knee replacement and varicose veins. The NHS is the first health care system in the world to routinely collect PROMs data and this could promote several initiatives such as informing patients, benchmarking performance, and linking payments to performance to name just a few [2].

The Patient Reported Outcome Measurement Group, Oxford, in their report to the Department of Health in 2010 [8] state there are three broad categories of PROMs: generic, preference based and disease specific. Generic instruments are intended to be used across many health conditions. Preference based instruments are also broad but can be used for cost-utility analysis. Disease specific are concerned with a specific disorder such as angina or a disease

specific to a population such as adolescents or a symptom such as loss of hearing. PROMs are not designed to elicit patient's satisfaction with their care. Black points out PROMs do not measure "*outcomes*" [20] but the patient's health "state" at a moment in time. Sequential use of PROMs can indicate a change in the patient's perception of their health over time, for example a year after a procedure. It is important to emphasise PROMs are not reporting on the *experience* of healthcare which could include how the patient was treated in an Outpatients Department.

The Department of Health in its *Guidance on the Routine Collection of Patient Reported Outcome Measures* [26] state that a PROM is a measure of a patient's health status or health-related quality of life. They go on to state they are typically short, self-completed questionnaires that measure the patient's health status or health related quality of life at a single point in time.

The Department of Health report *High Quality Care For All* [27] stated payments to hospitals will be conditional on the quality of care given to patients. The quality measures would include clinical outcomes and the patient experience and views about the success of their treatment. The report pointed out that the NHS budget in 1996/7 was £33 billion and in 2008/9 it was £96 billion. Obtaining value for money is important but evaluation of the effectiveness of interventions will include the patient's perspective. The report gives examples of the use of PROMs as improvement in pain-free movement after joint replacement or returning to work after treatment for depression.

1.5.1 Preference Based PROMs

Preference based PROMs are described as broad in content but also provide utilities or values regarding health [8]. In chapter 3, I will present a detailed review of the PROM instruments I will use in this thesis. The Patient Reported Outcome Group Oxford list several PROM instruments aimed at coronary revascularisation [8]. Their recommendation was the EuroQol Group EQ-5D instrument (EQ-5D) [28]. This instrument contains two separate elements. The first consists of five questions concerning mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each of the questions has five options and the patient selects one for each question. The options are "I have no problems", "I have slight problems", "I have moderate problems", "I have severe problems" and "I am unable to do my usual activity". The second part is a VAS scale from zero to one hundred. Patients are asked to place a cross on the scale where one hundred represents the best possible health and zero the worst.

1.5.2 Generic PROMs

Generic PROMs are intended to be relevant to a wide range of patient conditions and the general population [8].In the category of generic based instruments, the Oxford Group recommend the Short Form 36 (SF-36) [29]. This instrument consists of eight domains and thirty-five questions. The thirty-sixth question relates to health changes over a year. The domains are physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional and mental health.

1.5.3 Disease Specific PROMs

In the category of disease specific the recommendation was to use the Seattle Angina Questionnaire (SAQ) [30]. This instrument contains five domains and nineteen questions. The domains are physical limitation, angina stability, angina frequency, treatment satisfaction and disease perception.

1.6 Thesis

In terms of thesis novelty, I will investigate how PROMs are used in interventional cardiology and whether there is agreement between clinicians and patients on patients' quality of life. Bass et al pointed out that studies show clinician reported assessments tend to underestimate burden and severity of symptoms when compared with patient ratings [31]. Discordance in assessing patients' quality of life may be especially relevant in situations where symptoms are not directly observable by clinicians. This can include pain, fatigue, sexual dysfunction and emotional disorders [10].

Chang et al point out that the use of PROMs has expanded and this reflects an increasing recognition of what they call "*patient centredness*" in high quality health care [10]. They also point out that historically formal evaluations from the patients' perspective have been undervalued. In terms of the potential value of PROMs, Cheng et al also point out there are a number of applications of high-quality data from the perspective of the patient. These include assisting the patient in making informed decisions about their care and allowing their clinicians in monitoring the patients' perspective about their care [10]. The quality of healthcare can be monitored via PROM data and this in turn can

impact policy makers in decisions on reimbursement to hospitals. In clinical trials, PROMs can be used as a primary endpoint. There was an increase in the use of PROMs in oncology trials in the United States, rising from 27% in 2007 to 33% in 2013 [32].

1.7 Use of Stent or Surgery Trial Data

My thesis analyses data collected as part of the Stent or Surgery Trial (SoS) [33]. The SoS trial compared CABG with PCI. However, participants were all required to complete multiple PROMs at several timepoints. In conversation with one of the SoS trial Principal Investigators I was informed that the intention of collecting the PROMs data was for an in-depth analysis at a later stage. The intention was to compare different PROMs data by, for example, comparing generic versus disease specific. The objective was for the data analysis to be in the form of a higher degree for a student. However, the SoS Principal Investigator became involved in other high-profile research which caused delays in finding a student to undertake PROMs analysis.

The SoS trial is discussed in Chapter 2. SoS was a randomised controlled trial that compared percutaneous coronary intervention (PCI) with coronary artery bypass graft (CABG). The primary outcome measure was the rate of repeat revascularisation. The trial recruited patients from fifty-three centres in Europe and Canada and randomised nine-hundred and eighty-eight patients. Along with collecting clinical trial data, each patient was asked to complete several distinct PROMs at baseline, six and twelve-months.

Chapter 3 will discuss the various PROM instruments used. Chapter 4 investigates discordance in reporting quality of life by both patients and by their clinicians who estimated their patients' quality of life. I believe this analysis may be unique.

Chapter 5 investigates two PROMs, the Seattle Angina Questionnaire (SAQ) and the EQ-5D and examines the degree of agreement between the two. I compared one aspect of the EQ-5D, the VAS with each of the five domains of the SAQ. I also examined the magnitude of change over twelve-months.

Chapter 6 investigates correlation between a shorter, seven question version of the SAQ and the full nineteen question SAQ. A single summary score was calculated for both the short and full SAQ. I will compare the two scores at all three time-points using scatter plots. I will also compare the summary score for CABG and PCI patients again using a scatter plot.

Chapter 7 investigates the advantages and disadvantages of questionnaires as a method of collecting data. I will examine issues around financial considerations, practicability, speed, scale and anonymity. I also examine participant fatigue, dishonesty and interpretation. Finally, I examine bias in questionnaires.

The final chapter is a summary of this thesis with some key conclusions.

The next chapter will discuss the Stent or Surgery Trial.

Chapter 2 Stent or Surgery Trial

The Stent or Surgery Trial (SoS) [33] was a comparison between stent assisted PCI versus CABG for patients with multivessel disease coronary artery disease (CAD). PROMs data from the SoS trial forms the basis of this thesis. SoS was a multi-centre international trial in fifty-three centres in Europe and Canada. The primary outcome measure was the rate of repeat revascularisation. Nine-hundred and eighty-eight patients were randomised. There was a minimum of one-year follow-up for all patients. The results have been reported elsewhere [33]. In summary, 21% of patients in the PCI group required repeat revascularisation compared with 6% in the CABG group. There were 12 deaths in the PCI group and 4 in the CABG group.

2.1 SoS Protocol

All trial operators were required to perform optimal revascularisation following their local best practice. CABG could be performed either on or off-pump. Each trial centre could use stents of their choice. Any PCI could be performed as a single or two stage procedure, provided a two stage was completed inside twenty-eight days. All clinical and adverse events were followed from the point of randomisation. Follow-up visits were performed at six- and twelve-months post randomisation. All patients were followed up for at least one year. The primary outcome measure was the rate of repeat revascularisation. There were several secondary outcomes. These were death, q-wave myocardial infarction, symptoms of angina, cardiac medication requirements, all-cause mortality and left-ventricular function. Symptoms of angina were assessed using the Canadian

Cardiovascular Society classification (CCS). Left-ventricular function was assessed by echocardiography (ECHO) and was analysed independently. Repeat revascularisations were any coronary revascularisation performed after the index procedure. If the procedure was planned as two-stages, then a repeat revascularisation was defined as an intervention performed on a lesion treated during the first stage. Due to the nature of the intervention allocated by randomisation it was not possible to blind either the trial clinicians or patients to the allocation.

Recruitment took place between November 1996 and December 1999. Patients needed to be symptomatic and have multivessel CAD and should be suitable for either PCI or CABG in the opinion of a surgeon and interventionalist. Randomisation was centre-specific and block sizes of two, four and six were used. The primary outcome was analysed using Cox proportional hazard models. Chi squared was used to analyse the proportions of patients with no angina (CCS = 0) or any degree of angina (CCS > 0).

2.2 SoS Results

Analysis was on an intention to treat basis. The mean age of patients was 61 years and 79% were men. Most repeat revascularisations occurred within the first year following randomisation. Of those patients initially randomised to PCI, 9% crossed over to CABG. The number of deaths in the SoS trial was small at just 16 (4 CABG, 12 PCI group). However, it was stated that in the PCI group, there were eight deaths due to cancer and one in the CABG group. This was attributed to chance. All-cause mortality was reported and was also classified as cardiac,

other vascular, non-cardiovascular and unclassified. The low rate of death in the CABG group could be a product of the “favourable” coronary anatomy and surgical risk profile. There was no difference in the composite outcome of death or non-fatal Q-wave myocardial infarction. Severe angina, classified as CCS class 3 or 4, was present in 46% of patients. In the PCI group, 85% of patients received a single procedure rather than a two-stage procedure. In those patients who were allocated a PCI, 94% of the lesions attempted were successfully revascularised. Of those patients allocated to CABG, 3% were performed without cardiopulmonary bypass (off-pump). Of those patients allocated to PCI, 21% required additional revascularisation over a median period of two years. This compares to just 6% in the CABG group for the same period. Although 17% of patients randomised to PCI required revascularisation in the first year, this is a reduction when compared to the era of balloon angioplasty.

2.3 Trial Conclusions

The SoS trial concluded there was a “survival advantage” with CABG and the use of coronary stents reduced the need for revascularisation compared to previous studies. However, this was a time of significant emerging developments such as drug eluting stents (DES) and glycoprotein IIb/IIIa antiplatelet agents along with new oral antiplatelet agents such as clopidogrel. These were considered to have significant potential to reduce revascularisation rates. The SoS trial required patients to be suitable for revascularisation by either CABG or PCI. This limited the trial population to a minority of patients. There remained the possibility that individual cardiologists or surgeons may be unwilling to submit their patients to

randomisation. Centres did also report difficulty in obtaining consent because patients were required to accept two very different treatment options when one, CABG, was more invasive than the other. The trial authors also estimated that centres only randomised between 3-6% of eligible patients.

2.4 Other Trials in This Period

Several other trials in this period compared PCI to CABG. Serruys et al compared CABG and stenting (PCI) for the treatment of multi-vessel disease (ARTS Trial) [34]. They found at one year there was no significant difference between the two groups in terms of death, strokes or myocardial infarction, the primary clinical end point. They also found PCI is associated with a greater need for repeat revascularisation.

King et al in the EAST trial - in the era before stents were routinely used - found that PCI and CABG did not differ significantly in respect of their composite end-point (death, Q-wave myocardial infarction and a large ischemic defect) after three-years [35]. The authors did point out the clinical usefulness and long-term prognostic value of their composite end-point had not been established. They did not report on any quality-of-life measures in the trial. However, at three years there was a greater proportion of revascularization in the PCI group, and angina was also more frequent in the PCI group. Weintraub et al also reported on the EAST trial and included data on quality of life [36]. They reported almost two-thirds of patients reported good or very good health but that there was a "strong trend" for CABG patients to believe they had recovered completely compared to the PCI group.

The RITA trial [37] - also before stents were routinely used - found that recovery from CABG takes longer than PCI but CABG leads to less angina. However, there was no significant differences in risk of death or MI. The BARI trial [38] followed up patients for five years following either CABG or PCI. The trial found that patients who were randomised to PCI were not significantly compromised in terms of five-year survival by comparison to CABG patients although there were more revascularisations. However, if the patient was diabetic, five-year survival was significantly better for the CABG group of patients.

The SoS trial asked patients to complete several PROMs at three time-points. The time-points were baseline, six and twelve-months post-randomisation. The PROM instruments completed were the Short-Form 36 (SF-36) [39], the SAQ [30], the Cardiac Health Profile (CHP) which includes the Canadian Cardiac Society (CCS) [40] assessment of angina burden and EQ-5D [41] . There is a full discussion of PROM instruments in the next chapter.

2.5 SF-36

The SF-36 is categorized by the Patient-Reported Outcome Measurement Group, Oxford [8] as a generic instrument. This means it is applicable to many conditions or patient groups. There are thirty-five questions in eight separate domains into which data is grouped. A thirty-sixth question compares health with a previous year, making thirty-six questions in total. The domains are physical functioning; role physical; bodily pain; general health; vitality; social functioning; role emotional health and mental health. The questions have a categorical response and scoring uses a weighted scoring algorithm. Each domain is scored separately,

and the scores transformed into a range from zero to one hundred. Higher scores indicate better health. In the SoS trial, 947 patients completed the SF-36 at baseline, 916 at six-months and 919 at twelve-months.

2.6 Seattle Angina Questionnaire

The SAQ was developed by Spertus et al as a disease specific functional status measure [30]. It consists of nineteen questions in five domains. The domains are physical limitation; angina stability; angina frequency; treatment satisfaction and disease perception. Each question is assigned an ordinal value with higher scores indicating better function. Each domain is scored separately and transformed into a score between zero and one hundred. There is no 'global' score. The SAQ measures the domains over the previous four-weeks. In the SoS trial, 945 patients completed the SAQ at baseline, 918 at six-months and 917 at one-year post-randomisation.

2.7 Cardiac Health Profile

The Cardiac Health Profile is a disease specific questionnaire [42]. It consists of three parts. These are the degree of angina assessed using the CCS scale [40], quality of life (sixteen questions) and psychological cost-benefit (two questions). I will examine the use of CCS which was reported by patients and by clinicians reporting on their assessment of the same patient's angina burden in Chapter 4. In the SoS trial, 959 patients completed the questionnaire – including the CCS - at baseline, 921 at six-months and 930 at twelve-months post-randomisation.

2.8 EuroQol EQ-5D

The EQ-5D is a generic instrument consisting of five domains and a VAS. The EQ-5D was intended to be used in conjunction with disease specific instruments [8]. The five domains are anxiety/depression; mobility; pain/discomfort; self-care and usual activities. Societal valuations, which are country specific, were used to establish weights for the index score. The possible range is -0.59 to 1.00. A score of -0.59 is a state worse than death and 1.00 equates to the best possible health. The VAS is a single scale from zero to one hundred where zero is the worst and one hundred the best imaginable health. It was intended that both the index score and VAS would be reported. In the SoS trial, 886 patients completed the EQ-5D at baseline, 892 at six-months and 906 at twelve-months post-randomisation.

2.9 SoS One-year Follow-up

The one-year follow-up results from the SoS trial were published in 2003 and made specific reference to the SAQ data [43]. The trial found that scores for physical limitation, angina frequency and quality of life improved significantly at six-months for both PCI and CABG. The PCI group of patients did have a greater number of interventions. However, overall treatment satisfaction not differ significantly between the CABG and PCI groups[43].

There have been several publications following the original results from the SoS trial. These include Neuropsychological outcomes [44], the impact of age on outcomes [45] and the impact of acute coronary syndromes [46].

The next chapter discusses PROM instruments in more detail.

Chapter 3 PROM Instruments

The earliest randomised clinical trials - with a focus on the major diseases of that time - tended to report mortality as their primary outcome. Later, studies began to examine other adverse events, testing therapies with the potential to, for example, reduce the rate of stroke or myocardial infarction. With improvements in public health, general standards of medical care and the fact that new treatments now need to demonstrate incremental benefit over existing, proven therapies – the absolute gain with a new approach may be modest. This creates problems in trial design. To secure the statistical power required to demonstrate treatment differences with precision would require large numbers of outcome events. This in turn demands either large numbers of patients, or prolonged follow-up or a study population with a very high baseline risk. These factors increase the complexity, cost and timeframe of studies. There has been a move towards the development of composite outcomes. For example, the combined rate of death or stroke or myocardial infarction, MACE events. This is a means of increasing event rates and creating studies that can be performed in a timely manner, with a reasonable budget.

It is important to note however that the vast majority of studies report outcome in terms of the rate of subsequent, observed clinical events. The impact of these events, and importantly the treatments that are applied, on the overall well-being of patients has often been neglected. A desirable clinical outcome does not necessarily equate to an improvement in the patients' quality of life .

PROMs are a means for patients to report their symptoms, health related quality of life or satisfaction with treatment [47]. This reporting is usually performed independent of their clinician or other caregiver. This information can supplement clinical data such as mortality or specified clinical events such as the rate of strokes. PROMs can in turn help to inform clinical decisions, disease management or, on a large scale, policy in healthcare.

PROMs can measure a specific symptom such as pain or depression, knowledge of a condition, function such as limb movement, compliance to therapy or health related quality of life. The World Health Organisation (WHO) defines Quality of Life (WHOQoL) as "*an individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns. It is a broad ranging concept affected in a complex way by the person's physical health, psychological state, personal beliefs, social relationships and their relationship to salient features of their environment*" [48]. The WHOQoL emphasises the individual's perception of themselves and their views of their wellbeing. The WHO state that diabetes involves poor body regulation of blood glucose which is well understood. However, the effect of illness on the perception individuals have of their social relationships, working capacity and financial status has received little attention [48]. Quality of life could be said to address the patient's need for *fulfilment* as opposed to symptom relief.

The name patient reported *outcome* measures is a misnomer as the patient is not reporting 'outcomes' but is answering a series of questions relating to their

quality of life at the current time or over the last month in the case of some scoring systems [20].

PROMs can provide high quality quantitative data to examine the effects of an intervention on the patient's day to day life and their functioning in society. Data can be used to inform discussions between patients and their clinicians. These discussions may focus on the clinical benefits of continuing treatment as opposed to the patient's belief concerning the value to themselves. One less obvious use of PROM data can be to inform funding decisions made by Clinical Commissioning Groups (CCG). Clinical effectiveness of a procedure on its own is not necessarily sufficient to warrant its continued funding. In the United Kingdom the National Institute for Health and Clinical Excellence (NICE) uses the EQ-5D to calculate quality adjusted life years (QALY). The EQ-5D asks five questions about aspects of daily living and also contains a visual analogue scale. A QALY is calculated using two factors. These are length of life and quality of life [49]. Weinstein et al [50] point out there are several assumptions made in calculating QALYs. These can be summarised as: value is measured in terms of preference, preferences measured across individuals can be aggregated and used for the whole group and a QALY is the same for anyone.

Lord Darzi in his report in 2007 recommended PROMs should have a greater roles in the NHS [7]. As a result, from 2008, the Standard NHS Contract for Acute Services included a requirement to collect PROMs for four surgical procedures. These were hip or knee replacement, groin hernia and varicose vein procedures.

Hamilton et al reported on the determinants of patient satisfaction following lower limb joint replacement [51]. This study involved a large cohort and is a good example of the use of PROMs to influence healthcare delivery. Their study followed 4,709 patients over a four-year period in a single United Kingdom hospital involving multiple surgeons. The authors pointed out that although there was rapid access to surgery and fewer complications, satisfaction with services had declined over several years. As joint replacement was a high-volume service and was closely monitored, this was a good choice to examine this paradox. The trial used the Oxford Hip or Knee Score and Medical Outcomes Study Short Form 12 (SF-12) PROM. These were completed at baseline, six- and twelve-months post-surgery. The authors pointed out the concept of "satisfaction" is widely used in consumer marketing. They defined satisfaction as "*an attitude like judgement following an act based on a series of product-consumer interactions*" [52]. They go on to say satisfaction has been used as a healthcare performance indicator for surgery in the UK and Europe [53].

The authors also point out that outcome and satisfaction, although associated, are not the same metric. The authors make the statement "*PROM scores are useful tools for the assessment of clinical outcome in which they focus primarily on pain relief*" [51]. It could be that failure to meet "*optimistic expectations*" is associated with dissatisfaction following surgery. Five factors were identified that explain the patients' overall satisfaction following surgery. These were meeting preoperative expectations, satisfactory pain relief, the patients' subjective hospital experience, pre-operative physical status and finally twelve-month

physical status. They also found that factors such as patient age, gender and other comorbidities did not impact satisfaction.

3.1 Categories of PROMs

PROMs can be broadly categorised into three groups. These are preference based, generic health states, and condition or population specific instruments [8]. A generic instrument can be completed by a large range of people regardless of their medical condition, age or gender. As such generic instruments such as the short form 36 (SF-36) have been used in multiple trials in cardiology.

A preference-based instrument is broadly similar in that it can be applied to a range of conditions or population groups. However, the additional advantage is preference-based instruments can be used to calculate utilities which can then be used to calculate quality adjusted life years (QALY) for cost utility analysis. An example of such an instrument and one of the most commonly used is the European Quality of Life instrument EQ-5D.

Condition specific instruments focus on a particular condition or disease, a particular population, a symptom, or a function. An example is the SAQ. This nineteen question, five domain questionnaire was intended to measure both physical and emotional effects of coronary artery disease over the previous four weeks. An example of one PROM from each category will now be discussed. These three were selected because they are commonly used in clinical research and two - EQ-5D and SAQ - are discussed in detail in other chapters in this thesis.

3.1.1 Preference Based: EQ-5D

The EuroQol Group consists of a network of researchers and was established in 1987. The EuroQol Group describe the EQ-5D as "*a standardised measure of health status developed by the EuroQol Group in order to provide a simple, generic measure of health for clinical and economic appraisal*" [28]. The advantages of the EQ-5D are that it provides a descriptive profile and single index value for health status. This can be used in both health care and population health surveys. It can be completed by the participant themselves, by telephone or face to face interviews with healthcare professionals. It takes a few minutes to complete and is undemanding for participants.

The EQ-5D-3L was introduced in 1990. There were two pages, one consisting of five questions with multiple-choice answers. The second page contained a VAS. The five dimensions cover mobility, self-care, usual activities, pain and discomfort and anxiety and depression. There were three levels of response to each question. These are "I have no problems with…", "I have some problems with…" or "I have extreme problems with…". The VAS is a vertical scale from zero to one-hundred. The zero is labelled "Worst imaginable health state" and the other extreme is labelled "Best imaginable health state". The respondent is asked to place a cross on the scale that represents their health state at that moment in time. This represents the individual 'quantitative' measure by the respondent. This differs from the first part, the multiple-choice questions which eventually produce an index value weighted on a country basis.

It was intended that the EQ-5D would be supplemented by other disease specific instruments [8]. Weights or data-sets were established in specific countries which were based on societal valuations of health and these weights are used to calculate an index score. The data-sets include the United Kingdom (UK), several other European countries and the United States of America (USA). The score can be in a range from -0.59 to plus 1. The minus score represents a state worse than death and a score of one represents the best imaginable health state. An example of a state worse than death could be a persistent vegetative state.

In 2005 the EuroQol Group setup a Task Force to improve the sensitivity of the EQ-5D and reduce ceiling effects. This is when a high proportion of respondents score at the high-end of a scale making it difficult to discriminate among respondents [54]. If, preoperatively, a respondent scores at the high end of the scale there is little room for improvement postoperatively. Conversely, those who score at the low end of the scale there is little room for a lower score postoperatively. This makes it difficult to discriminate change at the extremes. The Task Force made no change to the number of dimensions, which remained at five. However, the number of levels of severity would be increased from three to five. The Task Force showed that five levels of severity would increase sensitivity (and the discriminatory power) whilst maintaining its feasibility and reducing ceiling effects [28]. The new levels would be: “I have no problems…”, “I have slight problems…”, “I have moderate problems…”, “I have severe problems…” and “I have extreme problems…”. As with the three-level response, each response represents a one-digit number from one to five. This results in a five-digit number ranging from 11111 (the lowest score) to 55555. These

numerals have no intrinsic arithmetic properties but are used to calculate a single index value.

The index value is calculated by using a "Crosswalk Index Value" calculator, provided by the EuroQol Group via their website https://euroqol.org/. The individual domain scores are pasted into the calculator, the correct country selected along with the total number of respondents. This then produces the EQ-5D profile (the five-digit number such as 12345) and the index value. This differs for each country. With five possible responses and five-levels, there are 3,125 possible health states. The best possible in nearly all countries is one, the exception being Zimbabwe where it is 0.9. The worst possible score, represented by 55555, is -0.594 for the UK but -0.109 for the USA. This shows that each country has a different expectation of what constitutes the worst or best imaginable health state. The EQ-5D can be used to calculate quality adjusted life years (QALYs). A QALY takes into account both quantity and quality of life [55]. The EQ-5D has been translated into over 170 languages [28].

3.1.2 Disease Specific: Seattle Angina Questionnaire

The SAQ was developed by Spertus et al in 1995 [30]. The SAQ measures both physical and emotional effects of coronary heart disease but only over the previous four weeks. The rationale for the development of the SAQ was that although clinical endpoints such as treadmill tests or left-ventricle function were important, they correlate poorly with the patient's functional status. Improving a patient's functional status is an important goal of any intervention. Spertus et al reasoned that quality of life domains should quantify the level of exertion as

patients tend to alter their level of activity to minimise their angina discomfort. They also wanted to measure the frequency of angina and the patient satisfaction with their treatment, alongside a functional measure to assess the patient's perception of how disease limits their life [30].

The SAQ measures nineteen items in five domains. These domains are physical limitation, angina stability, angina frequency, treatment satisfaction and disease perception (now known as quality of life). The largest number of questions appear in the physical limitation domain. Nine questions are grouped into three types of activity, from the lowest exertion such as dressing yourself, the medium exertion such as gardening to the highest exertion such as running or jogging. Activities were selected to minimise differences in socioeconomic groups. Each domain is scored separately, and the score transposed into a score from zero to one hundred. There is no overall single score.

In developing the SAQ the researchers studied four distinct patients groups. These were patients undergoing treadmill tests, outpatients with self-reported coronary artery disease, patients with initially stable coronary artery disease and patients undergoing percutaneous coronary angioplasty. They tested for validity, responsiveness and reproducibility. The authors suggested there were two primary advantages of the SAQ. The first was that it quantifies a broader range of disease effects than the CCS classification, the Duke Activity Status Index (DASI) or the Specific Activity Scale. The second was that the SAQ physical limitation scale captures activity that is specific to coronary artery disease unlike other instruments such as DASI or Short Form 36.

The SAQ has been used frequently in clinical trials over the years and used as a performance measure in assessing quality of care [56]. There are however several limiting factors in the use of the SAQ. One is its length at nineteen questions and the second is the lack of a single 'global' score. A single number would aid both physicians and patients' understanding of any changes over time. Another limitation is the original validation of the instrument involved mainly elderly males from a Veterans Affairs medical centre, and this may limit the generalisability of the SAQ. A more general limitation of any disease specific instrument, although not specifically the SAQ, is the effect of other comorbidities on the patients' responses. A patient with rheumatoid arthritis and coronary disease may have a flare up of their arthritis but no change in their CAD. This could be interpreted in the score as a worsening of their CAD. A shorter, seven question version of the SAQ has been developed as a response to the limitations mentioned above [57]and is discussed in detail in Chapter 6.

3.1.3 Generic: Short Form 36

The SF-36 is a generic instrument measuring both physical and mental health. The development of the SF-36 was driven by a recognition of the need to measure population health and not focus on a specific disease, condition or population group. It was developed to assess basic human values that are fundamental to everyone's functional status [39]. It does not target either a specific age group or disease but is intended for use by the general population. Its development was a result of recognition of the need to incorporate standardised health surveys in clinical trials and the development of general

population health surveys in the 1980s [39]. The SF-36 was preceded by the Medical Outcomes Study (MOS), which was a four-year longitudinal study of the variations in practice styles and health outcomes of chronically ill patients [58]. MOS measured forty health questions and involved twenty-three thousand patients from three-hundred and sixty-two clinicians and one-hundred and sixty-one mental health providers [39]. These in turn formed the basis for the shorter SF-36 which was first available in 1988.

The SF-36 records patients' responses in eight domains. These domains were chosen as they most accurately reflect disease and health conditions and those most widely measured by other health surveys [39]. The eight domains can be split into two groups of four. These are physical health and mental health. The four domains representing physical health are physical functioning, role physical, bodily pain and general health. The four domains representing mental health are vitality, social functioning, role emotional and mental health. There are ten questions that relate to physical functioning. In this domain the desire was to represent the extremes of activities, such as lifting, climbing stairs, kneeling or walking moderate distances. Role physical , four questions, examines health related role limitations. An example would be reductions in time spent in work or other usual activities. Bodily pain, two questions, looks at how pain interferes with work and its intensity. General health contains five questions and concentrates on the respondents view on their expectations. Vitality refers to levels of energy over four questions. Social functioning, two questions, focuses on both the quantity and quality of social functioning. Role emotional, three questions, examines happiness, stress and emotional problems. Mental health

asks five questions about the respondents feelings, such as feeling downhearted or being nervous. There is one question that asks about how the respondent's health compares to a year ago.

The SF-36 is scored in several stages [59]. First, groups of questions are assigned a value of between zero and one hundred, depending on the actual score of the question. For example, a score of 1 for individual items in the group of questions is assigned a score of 100. A high score denotes a more favourable health state. The second stage is to average the items together to create each individual scale. Blank fields are ignored.

Although the SF-36 is ubiquitous and familiar to many researchers there are a number of limitations. In a study of nursing home residents, Andresen et al [60] found skewed scores for some scales. This, they suggested, may limit the SF-36 to respondents with higher cognitive and physical functioning than typical nursing home residents. An instrument aimed at the general population and reporting on such a wide range of physical and mental health factors may not be responsive to a change in one aspect of the patients' condition, such as angina. If the patient already has multiple comorbidities such as severe arthritis and back pain, a small improvement in the angina burden following revascularisation may not be reflected in the SF-36 score. The 'signal' from angina pain is lost in the 'noise' of other chronic pain and discomfort.

3.1.4 PROMs Measure Different Aspects of QoL

It is important to understand that the PROMs reviewed above measure different aspects of a patient's QoL. EQ-5D is broad in its scope, applicable to a wide range

of the population and can be used to calculate QALYs. The two parts, the five domains and VAS are intended to be used together although they measure different aspects of the patient's QoL. The EQ-5D was intended to be used in combination with a disease specific instrument [8]. By contrast, the SAQ was developed to measure quality of life for patients experiencing angina. It is specific to a population with known heart disease. Unlike the EQ-5D, there is no country specific calculation of the domain scores. The SF-36 is generic and can be used in a large population of healthy people. It attempts to measure both physical and mental health and to reflect other health surveys [39].

The CCS scale for reporting angina symptoms is used in clinical trials [40] but is not strictly a PROM. It is predominately completed by physicians to rate their patient's functional status and is not generally completed by patients themselves. CCS is part of the Cardiac Health Profile which is a PROM.

The next chapter will discuss discordance in the reporting of patients' quality of life between patients' and their physicians.

Chapter 4 Discordance Between Physician and Patient

4.1 Background

This chapter examines discordance between patients and physicians in reporting of angina symptoms before and after revascularisation. Primary outcomes in studies examining revascularisation for coronary artery disease usually report the rates of occurrence of subsequent adverse events. These include mortality, subsequent myocardial infarction and the need for additional unplanned revascularisation [2].

There has been increasing recognition of the value of PROMs in the evaluation of treatment effect, symptom burden and quality of life. [8, 20, 26, 61]. There is, however, limited work comparing the reporting of functional improvement by patients and clinicians following coronary revascularisation.

Revascularisation procedures aim to restore or improve blood flow to the heart muscle and include CABG and PCI. Revascularisation can have a prognostic benefit in certain patterns of more advanced disease or when performed in the setting of an acute coronary syndrome. Most procedures, however, are performed to reduce angina symptoms [17, 62]

4.2 Literature Review

This chapter builds upon a research paper published by Kemp et al in 2019 [63]. A literature review for this chapter was further updated in July 2020. An iterative process was used to refine the search string. The final search string used was as follows: *"Patient physician discordance" OR "patient and physician reporting" OR*

"Patient and physician agreement" OR "Patient-reported symptoms and their documentation" OR "Patient and physician discordance" This returned twenty-one results. These were downloaded into a spreadsheet and each was then reviewed for relevance. This identified three articles which are discussed in detail. The remaining eighteen (Table 1) are discussed in summary form.

Lead author	Title	Cohort	PROM Used	Test	Author conclusions
Barton	Patient-physician discordance in assessments of global disease severity in rheumatoid arthritis.	223	Health Assessment Questionnaire, Patient Health Questionnaire (9 item) for depression, VAS for global disease severity	1 sample t-test	Nearly one-third of RA patients differed from their physicians to a meaningful degree in assessment of global disease severity. Higher depressive symptoms were associated with discordance. Further investigation of the relationships between mood, disease activity, and discordance may guide interventions to improve care for adults with RA.
Desthieux	Patient-physician discordance in global assessment in early spondyloarthritis and its change over time: the DESIR cohort.	702	BASDI (Bath Ankylosing Spondylitis Activity Index) and SF-36	Paired t-test	Discordance was not a stable trait, indicating discordance is not a patient characteristic.
Desthieux	Determinants of Patient-Physician Discordance in Global Assessment in Psoriatic Arthritis: A Multi-center European Study.	460	Health Assessment Questionnaire (HAS) and Disease Activity Index for Psoriatic Arthritis (DAPSA)	Univariable and multi-variable linear regression	Discordance concerned 29.1% of these patient/physician dyads, mainly by PGA>PhGA. Factors associated with discordance were psychological rather than physical domains of health. Discordance was more frequent in patients in remission, indicating more work is needed on the patient perspective regarding disease activity.

Lead author	Title	Cohort	PROM Used	Test	Author conclusions
Desthieux	Patient-Physician Discordance in Global Assessment in Rheumatoid Arthritis: A Systematic Literature Review With Meta-Analysis.	11879	VAS		Discordance in global assessment was most frequently defined as a difference of 3 points or more; even with such a stringent definition, up to half the patients were found to be discordant. The long-term consequences of this discordance remain to be determined.
Di maio	Symptomatic toxicities experienced during anticancer treatment: agreement between patient and physician reporting in three randomized trials.	1090	European Organisation for Research and Treatment of Cancer (EORTC) and QLQ-30	Cohen's kappa	Subjective toxicities are at high risk of under-reporting by physicians, even when prospectively collected within randomized trials. This strongly supports the incorporation of patient-reported outcomes into toxicity reporting in clinical trials.
Dobkin	Patient-physician discordance in fibromyalgia.	182	Patient-Physician Discordance Scale	Principal component analysis	The highest discordance score was on satisfaction with the office visit; physicians systematically underestimated patients' level of satisfaction.
Douglas	Patient-physician discordance in goals of care for patients with advanced cancer.	378	VAS	Chi sq Fishers exact	The data indicate the presence of significant ongoing oncologist-patient discordance with respect to goals of care. Early use of a simple visual analogue scale to assess goals of care can inform the oncologist about the patient's goals and lead to delivery of care that is aligned with patient goals.

Lead author	Title	Cohort	PROM Used	Test	Author conclusions
Fourrier-Reglat	When patients report diseases that prescribers seem unaware of: discordance between patient and physician reporting of risk-related previous history in NSAID users from the CADEUS study.	26618	Prescriber questionnaire	Multivariate analysis	The study showed that a substantial proportion of prescribers seemed unaware of the presence of risk related PMHs that the patient reported when asked.
Ghukasyan	Patient and physician agreement on reported Bath Ankylosing Spondylitis Disease Activity Index in patients with axial spondyloarthritis.	50	BASDAI	Interclass coefficient	Results could prompt the scientific community to revise outcomes or at least provide clear recommendations on the optimal way to collect it.
Henderson	Patient and physician reporting of symptoms and health-related quality of life in trials of treatment for early prostate cancer: considerations for future studies.		SF-12 recommended in review		Although novel therapies may produce different toxicities, the current recommendations should help to produce trial protocols that will allow comparable data to be generated within clinical trials comparing outcome from surgery, radiotherapy and other targeted ablative therapies for EPC.
Jacome	Patient-physician discordance in assessment of adherence to inhaled controller medication: a cross-sectional analysis of two cohorts.	395	VAS	Wilcoxon and Mann Whitney	Although both patients and physicians report high inhaler adherence, discordance occurred in half of cases. Implementation of objective adherence measures and effective communication are needed to improve patient-physician agreement.

Lead author	Title	Cohort	PROM Used	Test	Author conclusions
Sewitch	Psychosocial correlates of patient-physician discordance in inflammatory bowel disease.	200	10 item questionnaire		Increased physician awareness that psychologically distressed patients have difficulty processing of clinically relevant information may lead to improved doctor-patient communication during an office visit.
Sonnenberg	Personal View: Patient-physician discordance about benefits and risks in gastroenterology decision-making.				In instances of potential complications associated with risky medical interventions, patients may receive less medical therapy in exchange for more procedural safety.
Tago	Influence of large joint involvement on patient-physician discordance in global assessment of rheumatoid arthritis disease activity analyzed by a novel joint index.	12043	Novel joint indices	Multivariate logistic analysis	RA care providers should focus on pain and functional disability to decrease PGA-PhGA discordance. High disease activity and large joint involvement decreased PGA-PhGA discordance, indicating that the number and distribution of affected joints influenced the perception of disease activity by patients with RA and their physicians.

Lead author	Title	Cohort	PROM Used	Test	Author conclusions
Thomas	Patient and physician agreement on abdominal pain severity and need for opioid analgesia.	30	VAS	Cohen's kappa	Overall, patients and physicians agreed on the question of whether pain was sufficient to warrant opioids in 71 of 90 (78.9%) assessments; the corresponding kappa statistic of .57 indicated moderate agreement (P < .0001). These results, indicating that patients and physicians usually agree on whether opioids are warranted for abdominal pain, have important implications for further research on ED analgesia in this population.
Tory	Patient and physician discordance of global disease assessment in juvenile dermatomyositis: findings from the Childhood Arthritis & Rheumatology Research Alliance Legacy Registry.	639	Disease Activity Core Sets, Patient/Parent Global Activity Assessment Score (PF gVAS), Physician Global Activity Assessment Score (MD gVAS)	Chi sq, multivariate logistic regression	Discordance between PF and MD gVAS was common in this JDM cohort. Overall, higher PF rating was associated with poorer patient reported outcome (PRO) scores, while higher MD rating was associated with poorer objective measures. This suggests PF and MD assessments of gVAS may be measuring different aspects of disease, highlighting the importance of integrating PROs into clinical practice and research.
Wang (2019)	Factors associated with patient-physician discordance in a prospective cohort of patients with psoriatic arthritis: An Asian perspective.	142	SF-36		Increased age, higher fatigue levels, higher pain score and poorer mental health may explain underestimation of disease activity by physicians. Physicians' overestimation of disease activity may be explained by higher swollen joint counts.

Lead author	Title	Cohort	PROM Used	Test	Author conclusions
Wang (2018)	Factors associated with patient-physician discordance in a prospective cohort of patients with psoriatic arthritis: An Asian perspective.	142	Patient Global Assessment (PGA)using VAS scale, Physician Global Assessment (PhGA) 11 point numeric rating scale.	Spearman's correlations	Increased age, higher fatigue levels, higher pain score and poorer mental health may explain underestimation of disease activity by physicians. Physicians' overestimation of disease activity may be explained by higher swollen joint counts.

Table 1 Summary of eighteen articles derived from literature search

4.2.1 Pakhomov et al

Pakhomov et al [64] discussed agreement between patient reported symptoms and how they were documented in medical records. The main symptoms of interest were chest pain, chest pressure, shortness of breath/dyspnoea and cough. They used two sources of information. These were patient provided forms and the electronic medical record for each patient identified by the use of natural language processing (NLP), regardless of whether the patient was an inpatient or outpatient. This was a convenience sample of 121,891 patients. A convenience sample is when subjects are enrolled according to their availability and accessibility [65]. It is quick, inexpensive and convenient. With each of the symptoms, the researchers randomly selected 200 patients who marked their forms and another via NLP 200 who did not complete forms. All patients were at least 18 years old. They found the overall positive agreement for chest pain was .74, dyspnoea was .76 and for cough was .63. In their discussion they noted they did not possess sufficient data to determine if any discordance between symptom reporting by patients and clinicians has any significant clinical

consequences. Cohen's kappa is a measurement of interrater reliability, the degree of agreement between raters that takes into account chance agreement [24]. Cohen's kappa for chest pain was 0.52, dyspnoea 0.46 and for cough it was 0.38. They did point out their finding indicated what they called "*substantial discordance*" between patient reporting and care provider documentation on the symptoms [64]. As far as their limitations were concerned, they note the generalizability of the study depends on availability of electronic medical records, something that is becoming more common in the USA. The level of literacy or proficiency in English were potential variables to consider. Demographic characteristics may also be important.

4.2.2 Shafiq et al

Shafiq et al reported on patient and physician discordance in reporting of angina among stable coronary artery disease patients [66]. Their cohort was 1257 outpatients with coronary artery disease in 25 cardiology outpatient practices. The patients completed the SAQ just before their visit. One domain, for angina frequency categorized their angina over the last four weeks as none, daily/weekly or monthly. Following the outpatient visit, cardiologists then estimated the frequency of their patients' angina. Shafiq et al pointed out one of the primary goals of treating patients is to optimize patients' quality of life. This does require the physician to understand and report the presence and frequency of angina. This, they say, is challenging. They state that understanding the accuracy of physicians' assessments of patients' angina is a crucial step in improving angina recognition and treatment. They stated that if there was a low

correlation between patients' and physicians' assessment of the patients' angina this would support efforts to include PROMs in clinical care.

The study used the angina frequency score which they state correlated well with daily diaries of angina. The domain score is scaled between 0 and 100, a higher score indicating less angina. They created three groups, daily/weekly (score ≤60), monthly (score 61 to 99) and no angina (score = 100). Immediately following an outpatient visit the clinician completed a case report form to indicate whether the patient had experienced any chest pain within the last four weeks while being blinded to the patients' SAQ scores. In a secondary analysis the study reported on whether cardiologist estimated, or patient reported angina correlated better with health-related quality of life. The quality of life scale of the SAQ was used along with the VAS of the EQ-5D. The reason for this secondary analysis was to assess whether the patient reported, or cardiologist estimated angina correlated more closely with the patient reported quality of life score. After the outpatient visit the results showed that 67% of patients reported no angina over the previous four weeks, 25% reported monthly angina and 8% daily/weekly symptoms. The cardiologists' estimates for the same patients were 76%, 7% and 17% respectively.

There was a moderately strong association between the patients' reports of angina frequency and their health-related quality of life. However, the association with physicians' estimates were weaker. When patients reported having monthly angina cardiologists estimated that 46% of patients did not have any chest pain in the month. In cases where patients did report daily or weekly

angina, 26% were reported by their cardiologist to have no angina. The authors state they found a "*significant discordance*" between patients' reports of angina frequency and their cardiologists' estimate of the patients' symptom frequency. The authors state that even when American Medical Association Physicians Consortium for Performance Improvement and the National Quality Forum endorsed the SAQ as a performance measure it was not adopted in clinical practice [67]. The authors found "*modest agreement*" between patient reported and cardiologist estimated angina frequency with a Cohen's kappa of 0.48.

This study only examined angina frequency. When patients reported no angina, their cardiologist agreed 93% of the time. In those patients who reported either daily or weekly angina, 26% of their cardiologists noted no angina.

4.2.3 Sewitch et al

Sewitch et al reported on differences between patients' and physicians' health perceptions [68]. They point out that " *lower patient physician discordance is associated with positive health outcomes*" [69]. Another way to say this is that low discordance equals good agreement between patients and their physicians. They also state that comparing the findings of several studies is difficult for a number of reasons. These include discordance being based on different single items none of which are validated. Item scores may focus on different aspects of patient's health or treatment. Finally, there is a lack of a clear definition of discordance.

Their aim was to develop what they called "*a psychometrically sound and comprehensive measure of patient physician discordance*" [68]. The population for the study were patients with inflammatory bowel disease (IBD). This was chosen

because IBD has an unpredictable course with remissions and exacerbations. The doctor patient relationship may influence outcomes [70]. To develop a Patient Physician Discordance Scale (PPDS) they reviewed areas of discordance. The two strategies were to review empirical studies where agreement on a single item was discussed and to review agreement through statistics such as Cohen's kappa which measures interrater reliability [24]. Two domains were identified, relating to patient's health status and to outpatient visits. Items were then selected for each domain. A gastroenterologist and a health psychologist were asked to select the top ten items. A consensus was then reached. The ten items were abdominal pain, disease activity, physical limitation, psychological distress, emotional well-being, problem discussed, personal issues discussed, expectation of a prescription, expectation of testing and patient satisfaction.

Each of the ten items was measured independently using a VAS. Following the outpatient visit, both the patient and their physician were given the PPDS and asked to rate their perceptions. The VAS rating was compared for each item. The authors concluded they had developed a "*feasible*" means of measuring patient physician discordance in patients with IBD. They noted that physicians were less perceptive than patients in assessment of psychological distress as opposed to pain or disease severity. Overall discordance was statistically higher for patients who started seeing their physician during the previous year. The authors hypothesized that this may be due to the developing relationship in the first year. The study lacked a "gold standard" measure for seven of the items in the scale. The authors also noted the lack of generalizability of the finding as only ten physicians were enrolled. The PPDS was however designed for a wide range of

chronic diseases. They commented that further measurements of patient physician discordance may impact adherence with medication and health services.

4.3 Literature Review – Summary of 18 Articles

A summary of the remaining eighteen articles identified in the literature search are in table 1. Discordance was a factor in fifteen of the eighteen publications (83%). The size of the cohort ranged between 30 and 12,043.

In terms of statistical tests, T-tests, Chi squared and Cohen's kappa were mentioned in two cases (11%). Univariate or multivariate regression was mentioned in four cases (22%). Fisher's exact and Wilcoxon rank sum and Mann Whitney were each mentioned once.

The PROM used varied but the most common was the VAS scale which was mentioned in 6 cases (33%). The Health Assessment Questionnaire was mentioned twice.

The review below is split into two groups. The first are five articles most relevant to the theme of this thesis, namely patients' quality of life. The remaining thirteen are less relevant but still worthy of comment. These are not in any order or priority or relevance.

4.3.1 Brief details of five cases

Barton et al [71] reported on patient physician discordance in assessments of global disease severity in rheumatoid arthritis. Their cohort was 223 patients. They concluded that nearly a third of patients differed from their physicians to a

meaningful degree in the assessment of disease severity. They also found the more depressed the patient the greater the discordance between patient and physician.

Desthieux et al (2015) [72] examined discordance between patients and physicians in reporting of global assessment of disease activity in early spondyloarthritis. The assessment used a number scale from zero to ten. This was a French longitudinal multi-centre study with 702 patients at baseline. They used a linear mixed model. Both patients global assessment (PGA) and physician global assessment (PhGA) were compared. They point out that in rheumatoid arthritis (RA) patients assessments are based on a subjective perception of pain and functional incapacity, but PhGA focuses on inflammation. However, they state that in axial spondyloarthritis little is known about patient physician discordance. Using a ten-point scale, discordance was defined as a binary variable of ≥3 points. Three groups were identified, PGA < PhGA, PGA = PhGA and PGA > PhGA. They found that the absolute mean PGA values were always higher than the mean PhGA values. At baseline 71.2% of patients had a global rating within two points of their physicians' rating. Over time they found that the frequency of discordance was stable.

Dobkin et al [73] examined discordance between patients and physicians in fibromyalgia. Their cohort was 182 women. The authors used the Patient-Physician Discordance Scale (PPDS). The PPDS was a questionnaire developed by the authors in the context of a study of patients with inflammatory bowel disease. The PPDS is a VAS consisting of ten items derived from literature on

physician-patient agreement. The items relate to functioning, expectations of the outpatient visit, communication and patient satisfaction with the visit. The authors pointed out that with fibromyalgia patients often presented with vague non-specific symptoms such as body pain and stiffness in the morning. This lends itself to discordance between patients and their physicians. They used Principal Component Analysis. They found that the highest discordance was on satisfaction with the office visit where physicians systematically underestimated patients levels of satisfaction. They concluded there was a gap between what patients and their physicians experience during the clinic visit.

Douglas et al [23] reported on patient physician discordance in goals of care for patients with advanced cancer. Their cohort was 378 patients. They concluded that there was significant oncologist-patient discordance relating to goals of care. They felt the early use of a VAS would inform the oncologist about the patients' goals and aid the delivery of care.

Henderson et al [74] reported on health related quality of life in early prostate cancer and patient and physician reporting of symptoms. This was a review of published studies. The purpose was to review quality of life questionnaires and make recommendations for future trials. For health-related quality of life three PROMs were reviewed, the short form 36, Short Form 12 and Short Form 8 as these cover several domains such as physical, general health and mental health. The authors stated that as therapies themselves may produce symptoms, future studies should include a general health related quality of life questionnaire.

4.3.2 Remaining Thirteen Articles

De Maio et al [75] examined agreement between the patient and physician in symptomatic toxicity in anticancer treatment. Their cohort was 1090 patients, and they concluded that toxicities were at high risk of under reporting by physicians. They strongly supported the inclusion of patient reported outcomes into toxicity reporting in clinical trials.

Thomas et al [76] examined patient and physician agreement on abdominal pain severity and the need for opioid analgesia. Their cohort was thirty patients. They concluded physicians underrate patients' abdominal pain severity as assessed by VAS but are just as likely as patients to perceive that opioid analgesia is warranted. They reported "*moderate agreement between physicians and patients on the dichotomous query: does the abdominal pain warrant opioid analgesia*".

Tory et al [77] reported on patient physician discordance in disease assessment in juvenile dermatomyositis. Their cohort was 639 patients. They concluded that discordance was common and that this highlighted the importance of integrating patient reported outcomes into clinical practice and research.

Jacome et al [78] reported on patient physician discordance in the assessment of adherence to inhaled asthma controller medication. Their cohort was 395 patients. They found discordance in half of cases.

Desthieux et al 2016 [79] performed a meta-analysis involving 11,879 patients suffering from rheumatoid arthritis. The purpose was to discover the drivers of patient-physician discordance. Discordance was based on ≥3 points on a 10-point

scale of PGA and PhGA. They found the percentage of patients with discordance was 43%. PGA was usually higher than PhGA and the main drivers were pain and functional incapacity.

Sewitch et al [80] reported on patient physician discordance in inflammatory bowel disease with a cohort of 200 patients. They used a ten-item questionnaire. They concluded that greater physician awareness of distressed patients difficulty in processing clinically relevant information could lead to improved patient-doctor communication.

Ghukasyan et al [81] reported on patient and physician discordance in axial spondyloarthritis. Their cohort was 50 patients and they used a PROM called Bath Ankylosing Spondylitis Disease Activity Index (BASDAI). Their analysis used interclass coefficient. They concluded their outcomes could prompt the scientific community to revise outcomes or provide clear recommendations on how to collect outcomes.

Fourrier-Reglat et al [82] reported on discordance between physician and patients of previous medical history in use of Non-steroidal anti-inflammatory drugs (NSAIDs). A concern was patients reporting diseases that prescribers seemed unaware of in the patients' past medical history. The cohort of 26,618 patients from the CADEUS study and they used a prescriber questionnaire and multivariate analysis. Their conclusions were that a substantial proportion of prescribers seemed unaware of the presence of risk-related past medical histories that the patient reported when asked.

Tago et al [83] reported on rheumatoid arthritis and the discordance between patients and physicians using a novel joint index. This was a study carried out in Japan. The cohort was 12,043 adults. This cohort was held in a national database called 'NinJa'. Multivariate logistic and linear regression models were used. They found the number and distribution of affected joints influenced the perception of disease activity by patients and their physicians.

Wang et al 2019 [84] reported on patient physician discordance in patients with psoriatic arthritis in a multi-ethnic Asian population in Singapore. The cohort was 142 patients and the authors used generalised linear regression and univariable analysis. The PROM used was the Short Form 36. They concluded that increased age, higher fatigue levels, higher pain score and poorer mental health may explain underestimation of disease activity by physicians.

Wang et al 2018 [85] also reported a year earlier on factors associated with patient-physician discordance. The cohort was 298 patients with ankylosing spondylitis in two tertiary centres in Singapore. They used the Bath Ankylosing Spondylitis Functional Index (BASFI). Several factors were found to be associated with patient physician discordance. These included BASFI scores, biological factors and level of education. They concluded that global pain scores, lower educational level and what they called "*current biologics*" were associated with greater patient/physician discordance.

Desthieux et al 2017 [86] reported on the determinants of patient physician discordance in global assessments of psoriatic arthritis. The cohort was 460 patients and they used a zero to ten numeric scale. Discordance was a difference

of ≥3 points. Univariate and multivariable linear regression was used. Approximately a third of cases showed a discordance and discordance was more common in patients in remission. The authors concluded more work is needed on the patient perspective regarding disease activity.

Sonnenberg [87] reported on in a personal view on patient physician discordance about risks and benefits in gastroenterology decision making. He presented two scenarios and used threshold analysis to produce different results depending on values assigned to identical medical events. He reported that safety and therapy are the two references that determine patient and physician utility functions. The author stated that gastroenterologists are more concerned with safety and prepared to spend more healthcare resources on safety because the occurrence of complications will affect their professional status. The author stated patients may receive less medical therapy in exchange for more procedural safety.

Following a more recent literature review I also identified the CADENCE study [88]. This examined angina frequency in 2031 patients attending 207 Australian Primary Care physicians. Patients completed the SAQ detailing angina in five domains (frequency, recent change of symptoms, physical limitations, quality of life and satisfaction with current treatment). GPs completed the CCS and were asked if the patient's angina was 'optimally controlled'. The angina frequency domain of the SAQ was compared with GPs CCS. The primary endpoint was prevalence of weekly angina which was found in 29% of the cohort. There was a "*close relationship*" between the frequency of angina and patients' perceptions of their quality of life assessed by the SAQ physical limitation and quality of life

scores. There was a discordance between GPs and patients' assessment of angina control. The GPs reported 61% of patients had minimal angina-related impediment in physical activity, equivalent to the CCS class I. This included 12% of patients who reported daily angina. GPs also considered patients' angina to be optimally controlled in 80% of cases. The study concluded that physicians often underestimated the extent of angina and its impact on patients' health status. Physicians reported angina as being optimally controlled even in subjects reporting frequent symptoms. I would argue patients may confuse any non-cardiac chest pain with angina, thus overstating the prevalence of angina assessed by patients.

4.4 Data

The details of the Stent or Surgery Trial was discussed in chapter 2. The design paper was published in 1999 [89] and the main results in the Lancet in 2002 [90].

Angina symptoms were reported in terms of CCS class, by clinicians, at baseline and again at 6- and 12-months follow-up (see Table 3). Angina rating was performed in the context of a face-to-face visit, by clinical staff trained for this role as part of their trial responsibilities. The descriptors of the CCS scale were presented in the trial CRF as explanatory notes – presented on the facing page of relevant section of the document. Patients - at the same time-points - completed a number of PROM instruments including the Cardiac Health Profile (CHP) which includes the CCS question [40]. The CHP form also included the same descriptors of the CCS scale but translated into the local language.

This affords a unique opportunity to compare patient and clinician reporting of angina both before and after revascularisation.

4.5 Statistical Analysis and Methods

This chapter reports on a post-hoc analysis of data from the SoS trial. All analyses were performed using SPSS v 24, except for the weighted linear kappa tests which were performed using the tool at Vasser statistics site [91]. Continuous data have been reported as means and standard deviations or medians and inter-quartile range as appropriate. Comparative tests were two-sided and a p value of ≤ 0.05 was assumed to indicate significance. All analyses have been performed on an intention to treat basis.

Trial subjects were dichotomised into angina free (CCS = 0) or not. This was compared with McNemar's test. I calculated the 95% confidence interval for the magnitude of the difference both at baseline and at follow up.

The magnitude and direction of any difference between individual pairs of observations was quantified by subtracting the patient score from the clinician score. Descriptive statistics are presented as frequency histograms and the calculation of the mean score and standard error of the mean. The values were compared using a Wilcoxon signed-rank test.

I examined the rates of additional revascularisation in the year after symptom reporting - calculating the confidence intervals with the method of Clopper and Pearson and comparing the rate between groups using Fishers exact text.

4.6 Results

The main results of the SoS trial have been published [90]. Table 2 reproduces the baseline characteristics of the population, typical for revascularisation studies. There is a predominance of males and the mean age is just over 60 years. There were no important differences between the randomised groups.

Category	PCI (n = 488)	CABG (n = 500)
Men	390 (80%)	392 (78%)
Age (mean, SD years)	61 (9.2)	62(9.5)
Previous myocardial infarction	214 (44%)	234 (47%)
Previous cerebrovascular accident	5 (1%)	14 (3%)
Previous transient ischaemic attack	7 (1%)	11 (2%)
Previous peripheral vascular disease	31 (6%)	35 (7%)
Previous history of cardiovascular disease	235/487 (48%)	240/499 (48%)
Type I diabetes	19 (4%)	9 (2%)
Type II non-insulin dependent diabetes	49 (10%)	65 (13%)
Hypertension	212 (43%)	235 (47%)
Hyperlipidaemia	258 (53%)	251 (50%)
Current smoker	77 (16%)	72 (14%)
Ex-smoker	259 (53%)	286 (57%)
Left-ventricular ejection fraction (mean %)	57% *	57% †
Two-vessel disease	303 (62%)	262 (52%)
Three-vessel disease	183 (38%)	236 (47%)

Table 2 Baseline characteristics of patients randomised in the Stent or Surgery trial
** n = 398 at baseline, † n = 373 at baseline*

Figure 1, below, describes the trial conduct and patient numbers at baseline and follow-up. I present specific information on the number of individual CCS observations made by the clinician, by patients and the resulting number of paired observations at each time point and for the trial groups created at

randomisation. There were 912, 886 and 887 sets of paired observations at baseline, 6 months and 12 months respectively.

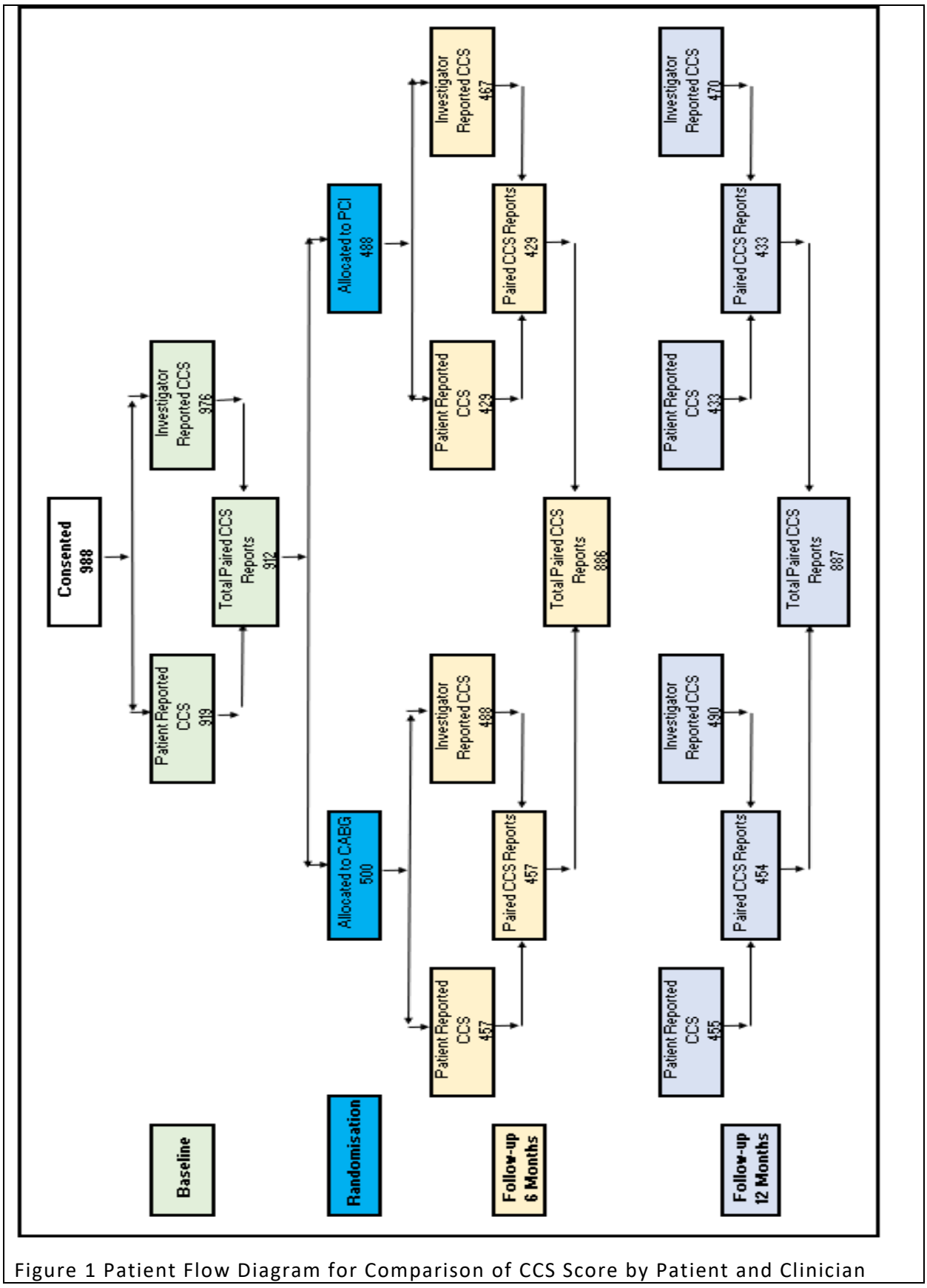


Figure 1 Patient Flow Diagram for Comparison of CCS Score by Patient and Clinician

Table 3 shows the number and proportion of subjects, at each time point, reported as manifesting each of the 5 possible CCS grades, with results for all recorded observations; for cases with paired clinician and patient information at that time point and, for these paired data, the information for the randomised treatment groups. From these data we can make some key observations. Clinician reporting is more complete than patient reporting. Our use of paired data does not result in substantial data loss, excluding about 6.5% and 7.6% of the clinician reported population at baseline and 12 months, respectively. The CCS group proportions are consistent between the individual clinician or patient reported and paired data sets, suggesting that the paired data information is representative of the whole study population.

Similarly, for all groups, a near identical distribution of proportions is seen at the 6- and 12-month follow-up points. Analyses at follow up were performed with the 12-month data but numerical results and conclusions would be representative of the findings at the earlier time point.

Baseline

CCS	ALL DATA (Unpaired)				ALL DATA (Paired)				CABG (Paired)				PCI (Paired)			
	Patient	%	Clinician	%	Patient	%	Clinician	%	Patient	%	Clinician	%	Patient	%	Clinician	%
	n=919		n=976		n=912		n=912		n=451		n=451		n=461		n=461	
0	72	7.83%	1	0.10%	70	7.7%	1	0.1%	29	6.4%	1	0.2%	41	8.9%	0	0.0%
1	187	20.35%	125	12.81%	184	20.2%	118	12.9%	92	20.4%	55	12.2%	92	20.0%	63	13.7%
2	285	31.01%	400	40.98%	284	31.1%	378	41.4%	141	31.3%	177	39.2%	143	31.0%	201	43.6%
3	215	23.39%	248	25.41%	214	23.5%	232	25.4%	107	23.7%	122	27.1%	107	23.2%	110	23.9%
4	160	17.41%	202	20.70%	160	17.5%	183	20.1%	82	18.2%	96	21.3%	78	16.9%	87	18.9%

Six Months

CCS	ALL DATA (Unpaired)				ALL DATA (Paired)				CABG (Paired)				PCI (Paired)			
	Patient	%	Clinician	%	Patient	%	Clinician	%	Patient	%	Clinician	%	Patient	%	Clinician	%
	n=886		n=955		n=886		n=886		n=457		n=457		n=429		n=429	
0	441	49.77%	677	70.89%	441	49.8%	635	71.7%	243	53.2%	374	81.8%	198	46.2%	261	60.8%
1	223	25.17%	181	18.95%	223	25.2%	165	18.6%	119	26.0%	66	14.4%	104	24.2%	99	23.1%
2	154	17.38%	73	7.64%	154	17.4%	67	7.6%	73	16.0%	16	3.5%	81	18.9%	51	11.9%
3	37	4.18%	20	2.09%	37	4.2%	15	1.7%	9	2.0%	1	0.2%	28	6.5%	14	3.3%
4	31	3.50%	4	0.42%	31	3.5%	4	0.5%	13	2.8%	0	0.0%	18	4.2%	4	0.9%

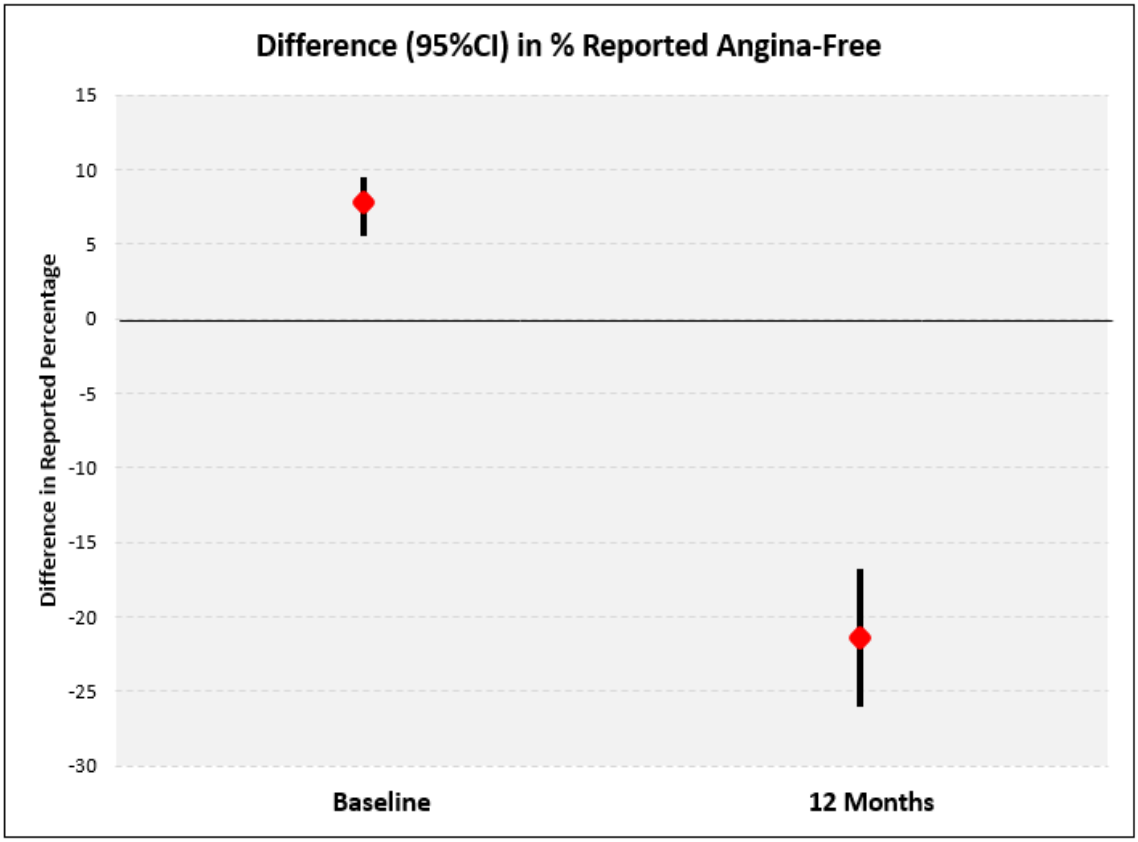
Twelve Months

CCS	ALL DATA (Unpaired)				ALL DATA (Paired)				CABG (Paired)				PCI (Paired)			
	Patient	%	Clinician	%	Patient	%	Clinician	%	Patient	%	Clinician	%	Patient	%	Clinician	%
	n=888		n=960		n=887		n=887		n=454		n=454		n=433		n=433	
0	449	50.56%	694	72.29%	449	50.6%	639	72.0%	261	57.5%	358	78.9%	188	43.4%	281	64.9%
1	209	23.54%	178	18.54%	209	23.6%	165	18.6%	92	20.3%	66	14.5%	117	27.0%	99	22.9%
2	160	18.02%	70	7.29%	159	17.9%	66	7.4%	71	15.6%	23	5.1%	88	20.3%	43	9.9%
3	42	4.73%	15	1.56%	42	4.7%	14	1.6%	18	4.0%	6	1.3%	24	5.5%	8	1.8%
4	28	3.15%	3	0.31%	28	3.2%	3	0.3%	12	2.6%	1	0.2%	16	3.7%	2	0.5%

Table 3 Proportion of subjects for each CCS grade as reported by patients and clinicians at baseline and follow-up

The proportion of subjects reported as being free from angina are summarised at Figure 2. The p-values are derived from McNemar's test. Physicians were reluctant to report freedom from angina at baseline, declaring CCS 0 in a single patient (1/912 = 0.01%). In contrast 70/912 = 7.7% of patients reported this status - Difference (95%CI) = 7.6% (5.8 – 9.3); p = < .001. At follow-up, the reverse was true with clinicians declaring 639/887 = 70.1% to be free of angina compared to 449/887 = 50.6% of patients - Difference (95%CI) = -21.4% (-17.1 - -25.8); p = < .001. Figure 2 also shows the separation of the confidence intervals

for the differences and confirms a substantial and significant change in the pattern of reporting from baseline to follow-up.



Time-point	Clinician Reported n/d = p%	Patient Reported n/d = p%	Delta (95% CI)	P value
Baseline	1/912 (0.1%)	70/912(7.7%)	7.6% (5.8 – 9.3)	.001*
12 Months	639/887 (72%)	449/887 (50.6%)	-21.4% (-17.1 - -25.8)	< .001*

* McNemar's test

Figure 2 Difference and associated 95% CI for subjects reported as angina-free by Clinicians and Patients at baseline and at 12 Months

4.6.1 Differences between clinician and patient gradings

Figure 3 shows frequency histograms for the difference in paired scores (clinician minus patient scores). At baseline there is agreement in just over a third of cases (36%). The linear weighted kappa statistic for overall agreement is 0.185. Cohen's kappa is discussed in detail in chapter 1, section 1.4.1. The distribution

of the observed differences is near normal suggesting a tendency of the clinicians to report more angina (38% of cases) rather than less (26% of cases).

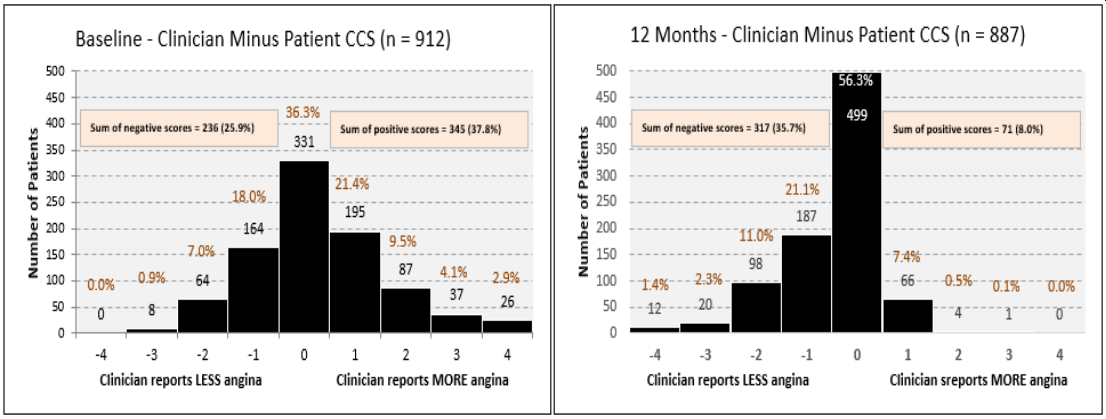


Figure 3 Baseline and 12 Months Clinician Minus Patient CCS Score, Paired Data

At follow-up, the distribution is very different. There is a greater proportion of paired values declaring the same grade (56% of cases). The weighted kappa for overall agreement is 0.312. The majority with discordant reporting now involves the clinician suggesting less angina rather than more (36% v 8% of cases).

I compared the distribution of the individual differences in the paired value reporting at baseline and 12 months using Wilcoxon Sign Rank test and found this difference to be significant p = <.001.

4.6.2 Subgroup analyses: Impact of Allocation, Patient Sex and Age

Figure 4, Figure 5 and Figure 6 show histograms relating to subgroup analyses for the treatment allocated at randomisation, patient sex and age, dichotomised at 65 years. The nature and magnitude of the differences between clinician and patient reporting is consistent across these subgroups.

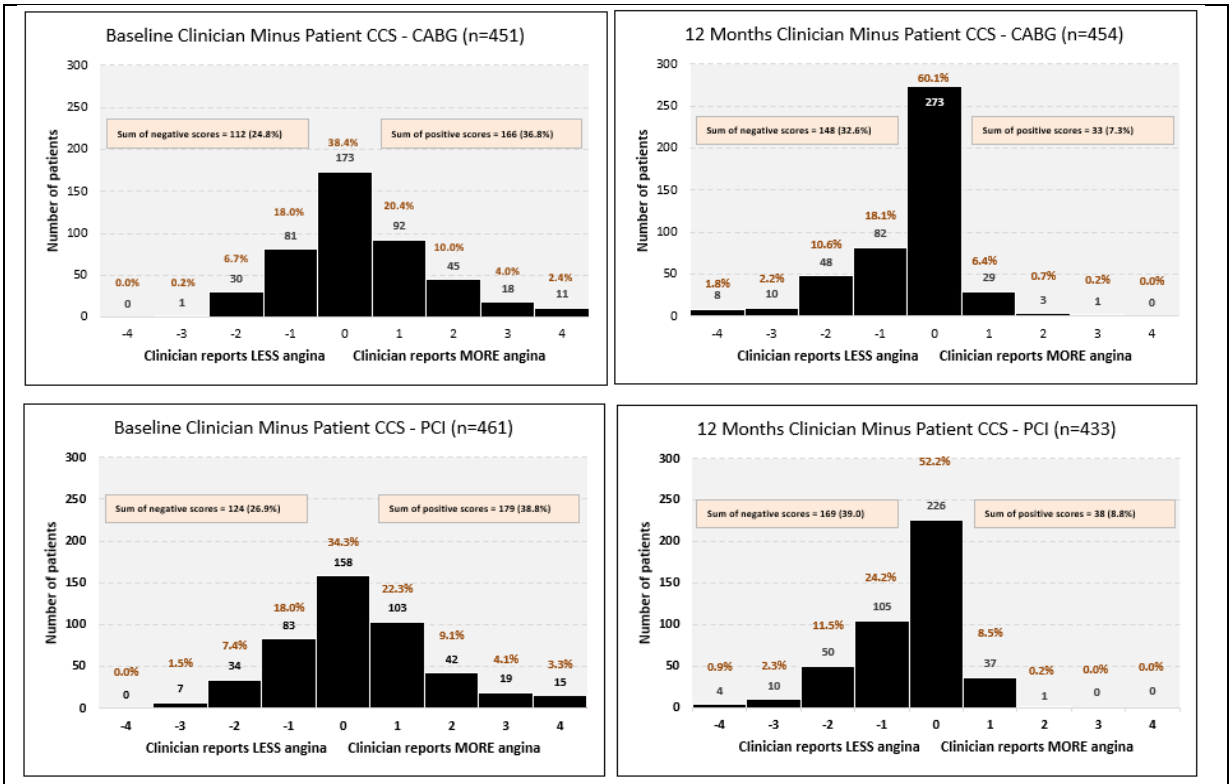


Figure 4 Baseline and 12M Clinician minus patient CCS – by allocation, Paired data

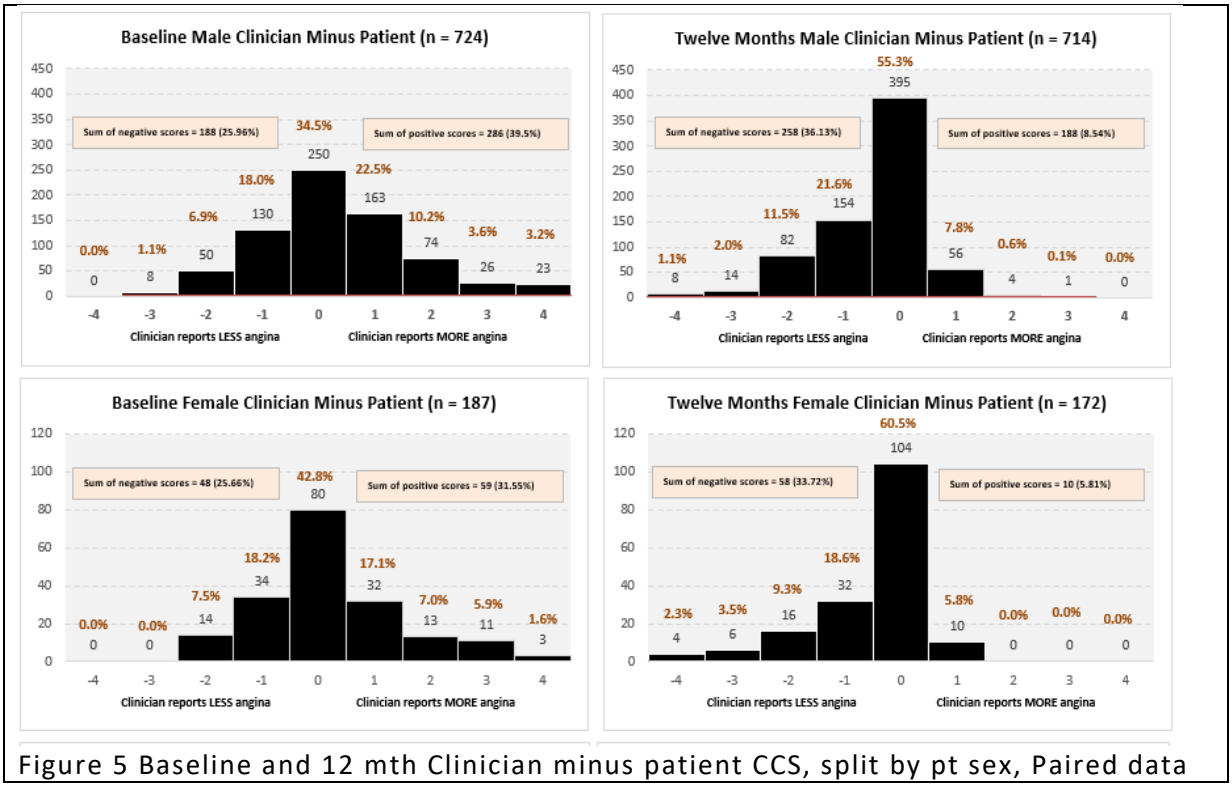


Figure 5 Baseline and 12 mth Clinician minus patient CCS, split by pt sex, Paired data

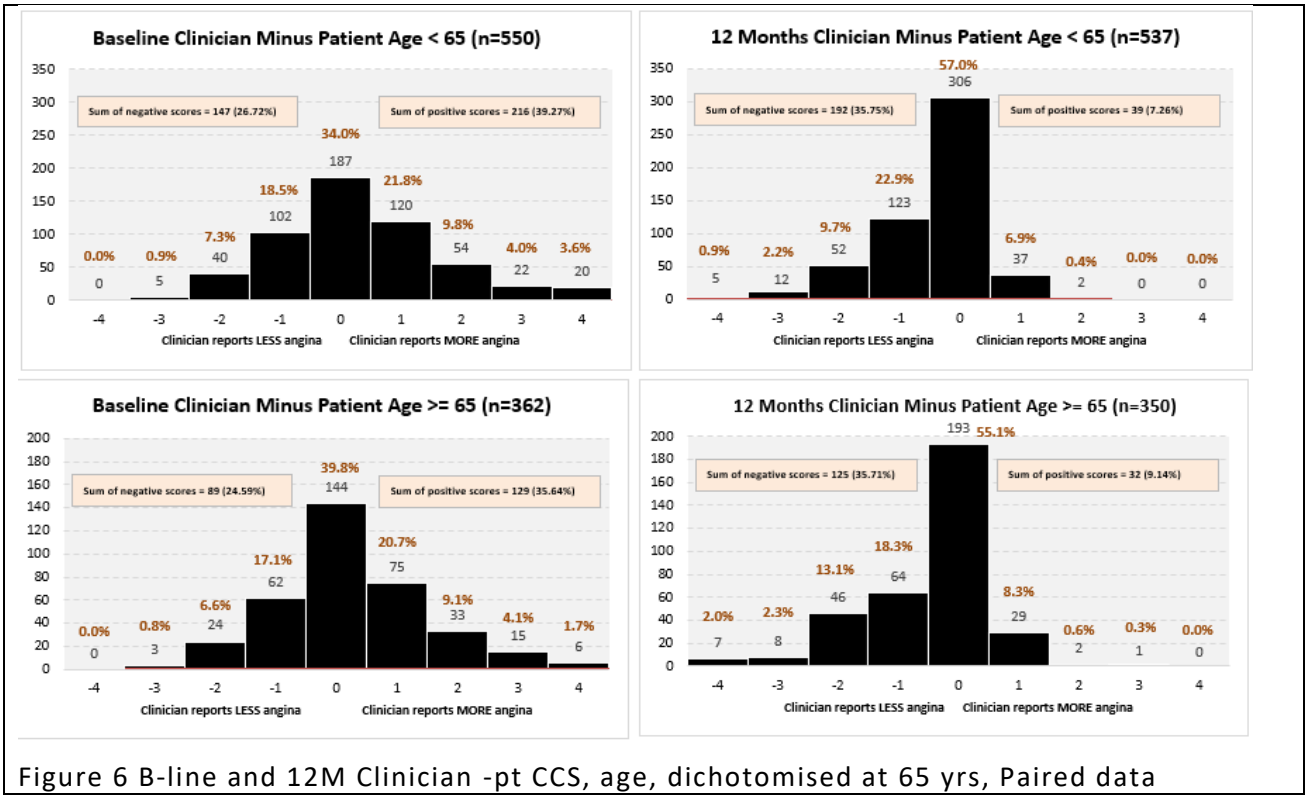


Figure 6 B-line and 12M Clinician -pt CCS, age, dichotomised at 65 yrs, Paired data

4.6.3 The impact of discordant symptom reporting on clinical outcome

I was interested if discordant reporting of angina at 12 months affected mortality and repeat revascularisation over the subsequent 12 months. Table 4 summarises the results. There are too few deaths to compare mortality between the groups. There were more additional revascularisation events, but the absolute numbers are modest and the associated confidence intervals are wide. I note however that, when clinicians report more angina at 12 months, the rate of revascularisation over the subsequent 12 months is higher than for trial subjects with concordant reporting: 7.04% versus 2%; P = 0.036.

Angina Reporting at 12 Months		Mortality	Revascularisation		
Angina Reporting Group	**n**	**n (%)**	**n (%)**	**% (95%CI)**	**P Value**
Agreement Between Clinician and Patient	499	5 (1%)	10 (2%)	2% (0.097 - 3.65)	
Clinician Reports More Angina	71	0 (0%)	5 (7.04%)	7.04% (2.33 - 15.67)	0.036
Patient Reports More Angina	317	2 (0.63%)	10 (3.15%)	3.15% (1.52 - 5.72)	0.36

The 95% CI has been calculated by the exact method of Clopper and Pearson.
The P values are from Fishers exact test comparing each of the groups with discordant reporting, with the concordant group

Table 4 Clinical events between 12 and 24 months in patients with concordant and discordant reporting of angina by clinicians and patients

4.7 Discussion

It was notable that the literature review in this chapter only identified twenty-one results. However, I was not able to identify a single publication comparing CCS scores reported by patients and assessed by clinicians. The review identified comparisons and discordance in diverse disease groups such as rheumatoid arthritis, cancer, NSAIDS and inflammatory bowel disease. Shafiq et al [(66)] did report on discordance of reporting angina pain between patients and physicians. However, their instrument was the SAQ, not CCS.

The results of my analysis of CCS scores suggest important differences in the reporting of angina by clinicians and patients, particularly when considered in relation to the timing of the observation - before and after revascularisation. At baseline there is reasonable agreement with a modest over-statement of angina by clinicians. At follow-up this is reversed with clinicians declaring a greater treatment effect than their patients. These findings have clear and important implications for our perception of previous research in this field which has, in the main, focussed on clinician reporting.

The CCS was first described in the literature in 1976 [40] and is a classification of symptom burden and has been used to evaluate patients angina burden (Table 5). CCS provides distinct grades, from one – the lowest - to four, describing the level of exertion that will induce angina. A grade zero is used to indicate no angina symptoms. CCS has been adopted worldwide [92].

Grade	Description
I	Ordinary physical activity does not cause angina, such as walking and climbing stairs. Angina with prolonged exertion at work or recreation.
II	Slight limitation of ordinary activity. Walking or climbing stairs rapidly, walking uphill, walking or stair climbing after meals, or in cold, or in wind, or under emotional stress, or only in the few hours after awakening. Walking more than two blocks on the level and climbing more than one flight of ordinary stairs at a normal pace and in normal conditions.
III	Marked limitation of ordinary physical activity. Walking one or two blocks on the level and climbing one flight of stairs in normal conditions and at normal pace.
IV	Inability to carry on any physical activity without discomfort, anginal syndrome may be present at rest.

Table 5 Canadian Cardiovascular Society grading of angina pectoris

I am not aware of any publications that compared clinicians' assessment of patients' angina burden with the patient's own assessment using the CCS instrument.

As detailed in the literature review, there are multiple instances of discordance between patients and physicians in reporting the patients' health related quality of life. A number of authors recommended the use of PROMs in clinical care. Pakhomov et al [64] found substantial discordance between patients and

physicians but were unclear if this had any effect on clinical outcomes. Shafiq et al [66] found low correlation between patients and physicians of the patients angina and they stated this supports the use of PROMs in clinical care. Sewitch et al [68] looked at perceptions in healthcare. They found physicians were less perceptive about psychological distress than of pain or the severity of the disease. Douglas et al [23] felt early use of a VAS would help to inform the oncologist about patient goals and aid the delivery of care. Tory et al [77] highlighted the importance of integrating PROMs into clinical practice and research. Sewitch et al [80] concluded greater physician awareness of distressed patients could lead to improved patient-doctor communication.

In the SoS trial, one inclusion criterion demanded angina symptoms and hence clinicians would be reluctant to declare CCS 0 as this would be a protocol violation. It is also possible that patients with symptoms at screening had been prescribed additional medical therapy that had taken effect before functional status was reported. Indeed, it is common for patients to have been prescribed medication for the relief of angina pain prior to a revascularisation procedure.

At follow-up patients may report persisting chest discomfort or related symptoms (such as breathlessness) as angina even if the characteristics of the symptoms appear to have a non-cardiac cause when assessed by the clinician. This may suggest a role for more objective evaluation, with exercise testing for symptoms and imaging evaluation for ischaemia. A recent study by Stone et al used formal adjudication for symptom reporting and, interestingly reported anginal rates comparable to those noted in SoS [93].

Clinicians may wish to be positive about the results of revascularisation and this may influence their reporting of angina symptoms resulting in a subconscious minimisation of the true symptom burden. Bias has been defined by Gludd as systematic errors that encourage one outcome over others [(94)]. In the SoS trial only a small proportion of the population were randomised. This could lead to selection bias if the 'healthier' patients were approach for consent. Blinding in the SoS trial was not possible due to the nature of the interventions, that is, either CABG or PCI. Blinding refers to participants such as patients, health-care providers and others being unaware of the random allocation [(94)].

Other potential sources of bias are more subtle. These can include cultural stereotyping of social groups, such as blacks or women [(95)]. Implicit bias is without conscious awareness and may impact clinical decision making and perpetuate health inequalities [(95)].

The nature of the consultation process may be suboptimal in terms of setting, time available or in the communication process such that the clinicians do not acquire an accurate perception of the true symptom state. Fischer and Ereaut explored five main themes around the patient clinician consultation [(96)]. These were making sense of the consultation itself; fear in the dynamic; invisible structures interfering with the consultation; the fragmented nature of conversations and the system itself. If the purpose of the consultation is not clear to both parties this can lead to a sub-optimal exchange. Fischer and Ereaut suggest that fear is a significant element in the consultation, from both the patient and clinician. The patient may worry about the diagnosis or not

mentioning their symptoms in part or at all. The clinician may fear missing a diagnosis or may be anxious about the interaction itself. Invisible structures relates to the way a clinician has a mind map of how the consultation should evolve. There is limited time and many patients, but the patient is unaware of the clinician's mind map. The invisible structures extend to the needs of the health administrators and other healthcare professionals. Everyone has their own idea about the 'system' works. Fragmented conversations refers to the hidden rules we all learn and employ in day-to-day conversation, such as taking turns. However, patient-clinician consultations are special cases with a different set of rules. One issue is that the patient may have several consultations over a year and may see this as one continuous exchange. However, the clinician may see hundreds of patient over the same time period and cannot reasonably be expected to recall the exact stage of each. Finally, system dynamics refers to the way a system evolved and can adapt. Fischer and Ereaut [(96)] suggest the quality of the consultation is reduced by the way the system has evolved and ability to innovate. Evolution has been bound up in the protected doctor and disempowered patient.

4.8 Limitations

These data are derived from a study conducted about 20 years ago and I cannot be sure that the results would translate to contemporary practice. The specific setting of a clinical trial may have affected reporting as, for example, clinicians - keen to recruit for the study - may have been inclined to report symptoms as this was an inclusion criterion.

The SoS trial did not involve any protocol directed, symptom-limited exercise testing or other objective tests for ischaemia. This made it difficult to identify the aetiology of chest discomfort reported at follow-up.

The analyses are restricted to trial subjects with angina reporting by both clinicians and patient, resulting in the exclusion of some of the randomised population. It is reassuring however that the requirement for pairing excluded only about 7% of the original population and that the distribution of reported CCS grades are very similar in the original and study datasets. Patient retention over the 12-month follow-up period was good with 16 patients dying and a further 16 lost to follow-up.

This study uses the CCS classification and limitations of this system have been identified in the literature [97, 98]. There are only 4 grades for the description of symptoms and a single grade may describe different types of limitation with an assumption of equivalence. There is only a weak relationship between the scale and anatomic disease or prognosis. I am not aware of any studies that have validated the use of the CCS grading system by patients and I cannot be sure that there will have been a consistent understanding of the explanatory text used in the presentation.

Mortality and revascularisation data was available to me as detailed in Table 4. However, I did not have access to broader morbidity data by random allocation. With full access to more detailed patient data a more nuanced analysis would have been possible.

4.9 Conclusions

The SOS trial was one of the first to report angina status by clinicians and patients in the same terms, both before and after coronary revascularisation. The results suggest that, when compared to patient reporting, clinicians may declare a modest overstatement of angina at baseline. In contrast, at follow-up, clinicians report a greater proportion of patients to be angina-free and tend to minimise the extent of symptoms in other subjects.

It is possible that studies reporting outcomes declared by clinicians may exaggerate the therapeutic effect as perceived by patients. This may happen if the clinicians fail to correctly estimate the effect of the intervention on the patients' quality of life. It may also be the case that patients overestimate the effectiveness of an intervention on their quality of life. Patients may be focused more on functional improvement rather than preventing clinical deterioration.

This chapter emphasises the importance of including patient reported outcomes in evaluating the treatment of coronary artery disease.

The next chapter examines the correlation between the VAS score and each domain of the SAQ.

Chapter 5 SAQ and EQ-5D VAS comparison

As part of the data collection process for the SoS trial, participants were asked to complete several PROM instruments. These included the SAQ, a disease specific PROM and the EQ-5D, a generic PROM. These instruments were completed at three time-points: baseline, six and twelve-months post randomisation. In this chapter, results from two PROMs will be compared to identify the consistency of scores. The first question is to what extent is there correlation between the individual domains of the SAQ and the EQ-5D VAS score. Baseline and twelve months data will be analysed as previous work has shown that there is little difference between six and twelve-months [63]. A second question is to what extent the instruments agree in reporting the direction and magnitude of change from baseline to follow-up at 12 months.

5.1 SAQ Domains

The SAQ measures both physical and emotional effects of ischaemic heart disease (or angina) over the previous four weeks [8]. There are five domains and nineteen questions. The five domains are physical limitation (PL) with nine questions; angina stability (AS) with one question; angina frequency (AF) with two questions; treatment satisfaction (TS) with four questions and quality of life (QL) with three questions. The angina frequency domain was included as this was considered both a critical measure of angina classification and also a prognostic indicator [30]. Treatment satisfaction was included as the responses to this question may influence the therapeutic strategy. The response to each question is assigned an ordinal value and each domain score is summed separately and

converted into a score between zero and a hundred. Higher scores indicate better function. To achieve this, I subtracted the lowest possible score from the actual score. I then divided by the range and multiplied by one-hundred [30]. I repeated this process for each domain.

The patient burden in completing the SAQ is low and the measure can take between five and ten minutes to complete [8]. Spertus et al suggest a change in score over time of ten points represented a change perceptible to patients and this was considered a clinically important difference in scores [30].

5.2 Method for Scoring EQ-5D

The EQ-5D used in the SoS trial has five dimensions of perceived problems for the respondent to answer. The five dimensions relate to mobility, self-care, usual activities, pain/ discomfort and anxiety / depression [41]. The EQ-5D used by the SoS trial contained three possible answers for each dimension. These range from level 1, "I have no problems…" to level 3, "I have extreme problems with…". A 'one' is the lowest score and three the highest. The response pattern for each questionnaire can, therefore, be expressed as five numbers. These range from 11111 for the best possible health state or 33333 for the worst. The EQ-5D can be used to calculate QALYs. To achieve this, the five digits representing the health states need to be converted into a single summary number, or index value. The index values are generated by using a 'Crosswalk Index Value Calculator' which is obtained from the EuroQol group (https://euroqol.org/). The calculator applies weights derived separately from several countries, including the United Kingdom (UK).

Following completion of the five dimensions the respondent is asked to complete a VAS. This is a vertical scale in the form of a 'ruler'. The scale is from zero at the bottom to one-hundred at the top. The respondent is asked to place an 'X' at the point on the scale which represents how they rate their health on that day. The EQ-5D reporting recommendations are that both the score derived from the five questions and the VAS are reported. There is no calculation of any correlation between the two scores. The VAS represents just the respondents' view of their health at that moment in time.

There are two reasons why I decided not to use the index value as a comparison against the individual SAQ domain scores. The crosswalk index value calculator is based on a specific country and the SoS trial was international in scope. Approximately a third of all randomised patients were residents of the UK. At the time, the crosswalk index calculator I used did not have values for nine countries participating in the SoS trial. This amounted to 415 randomised patients at baseline from a total of 988 (42%). Omitting these countries data could have introduced bias.

The second reason why I decided to not use the index value is related to the first. In general, the EQ-5D is calibrated for utility / cost analysis. The index value is based on a country average rather than being 'person specific'. The index value for a patient represents how 'good' or 'bad' the health state is according to the preferences of the general population of a country. Therefore, the index value for a patient is "changed" to some extent and reflects the values of a country. As the SAQ domain score is based on the patients' responses, I wanted to compare

this to the VAS which is asking the individual patient how they feel at that moment in time.

5.3 Mean Imputation of Missing Data

When the EQ-5D data was examined, there were twenty-four cases at baseline (Figure 7) and sixteen at twelve-months (Figure 8) where only four of the five domains were answered. This represents 2.71% rate of imputation at baseline from 886 cases where there was a date of completion of the EQ-5D and 1.76% imputation at twelve-months. In these cases, I decided to impute the value by mean substitution, calculating the mean of the other four values and rounding this to a whole number.

BL_Q1	BL_Q2	BL_Q3	BL_Q4	BL_Q5	Imputed Field
2	1	2	2	1	Q3
1	1	1	1	1	Q4
1	1	1	1	1	Q4
1	1	1	1	2	Q1
2	1	2	2	2	Q1
1	1	1	2	1	Q3
1	1	2	1	1	Q4
2	2	2	2	1	Q2
2	1	2	2	2	Q5
2	1	3	2	1	Q1
2	1	3	2	2	Q4
2	1	2	2	2	Q5
2	1	2	2	1	Q3
2	2	2	2	2	Q2
1	1	1	1	2	Q4
1	1	1	1	2	Q4
1	1	2	2	2	Q4
1	2	2	3	2	Q2
1	1	1	2	1	Q1
2	1	2	2	2	Q4
1	1	1	2	1	Q5
1	1	1	2	1	Q1
1	1	1	1	1	Q5
2	2	2	2	3	Q2

Figure 7 Details of 24 imputed values for EQ-5D at baseline
BL = baseline, Q1 = Question 1

M12_Q1	M12_Q2	M12_Q3	M12_Q4	M12_Q5	Imputed Field
2	1	2	1	2	Q5
2	2	3	2	2	Q2
2	1	2	2	2	Q1
1	1	1	1	1	Q5
1	1	1	1	1	Q2
1	1	1	1	2	Q2
2	1	1	3	2	Q5
1	1	2	2	2	Q5
2	1	1	1	1	Q4
2	2	2	2	2	Q5
2	2	2	2	2	Q2
1	1	1	1	2	Q4
1	1	1	2	3	Q4
1	1	1	1	1	Q4
2	1	2	2	2	Q5
1	1	1	1	1	Q2

Figure 8 Details of 16 imputed values for EQ-5D at 12 months
M12 = Month twelve, Q1 = Question 1

Rombach et al [99] examined multiple imputation for PROMs. They found that imputation at item or overall score level provided similar results with sample sizes greater than five hundred. Simons et al [100] examined imputation to deal with missing EQ-5D data. They also found sample size to be important, with little difference in results with less than 500 records and missing data of between 5 and 10%. They found index imputation to be more accurate with sample sizes of over five hundred.

The SAQ data also contained missing values for each domain. Missing values were identified and dealt with separately for each domain using mean substitution. Mean substitution is recommended if there is a response to at least 51% of the questions in a domain [101]. Only three of the five domains contained three or more questions, allowing mean substitution. In the physical limitation

domain, imputation was used in cases where there were up to four fields with no data for the time-point. In the domains of treatment satisfaction and disease perception only cases with one question where data was missing used mean substitution. The details of the number of records and the count of cases where mean substitution was used at baseline and twelve-months are detailed in Table 6.

Time-point	Total Respondents	Physical Limitation 9 Questions Cases Imputed (%)	Treatment Satisfaction 4 Questions Cases Imputed	Disease Perception 3 Questions Cases Imputed
Baseline	945	57 (6%)	5 (0.5%)	1 (0.1%)
12 Months	883	58 (6.6%)	32 (3.6%)	38 (4.3%)

Table 6 SAQ count of respondents and imputed values across two time-points

5.4 Method of calculating differences between VAS and SAQ domains score

The SoS database was accessed and a query created to extract data relating to both the EQ-5D VAS and the SAQ domains. This data was in relation to baseline and twelve-months. The query was output as an Excel spreadsheet. The columns were time point, VAS and five columns relating to the individual SAQ domains. I then created five more columns relating to the VAS minus the individual domain score.

The second question concerning the sensitivity to change required another spreadsheet. This contained columns for VAS and each domain score at baseline and twelve-months on the same row for each patient. A column was created for the VAS at twelve-months minus the baseline VAS. Similar columns were then created for each SAQ domain score. It was then possible to import this into SPSS.

As you would expect the patients' quality of life to improve after revascularisation the value for change should be a plus figure in most cases.

5.5 Statistical Tests

The data was analysed in SPSS v 26 in August 2020. I examined the relationship between the scores from the different SAQ domains and the EQ-5D VAS as reported by the same patient at simultaneous time points. Frequency histograms were created for the VAS and SAQ domains at baseline and again at twelve-months. I also report the number of cases, mean and standard deviation. Interquartile range (IQR) is reported for each of the SAQ domains and VAS. Scatter plots were created which included the correlation coefficient and the regression equation for the best fit line. Five Bland Altman plots, one for each SAQ domain, are created for baseline and another five for twelve-months. For the test of sensitivity to change from baseline to twelve-months, Wilcoxon rank sum was used to compare the VAS change to each separate SAQ domain change.

5.6 Statistics

The VAS score at baseline is a range between 0 and 100. Table 7 only displays the VAS scores at baseline and twelve-months that are divisible by 5, for example, 5 or 10 etc. At baseline, of a population of 819 cases, 80% are divisible by 5 and 57% divisible by 10. At twelve-months the figures for a population of 860 were 73% and 48% respectively. All the baseline VAS scores between 0 and 50 that were *not* divisible by 5 amounted to 69 cases, or 8% of the population. This demonstrates a propensity for respondents to select a rounded figure relating to their VAS score [(102)]. The same issue did not occur for the SAQ, as shown in Table

8. Jain et al [103] found that people find non round numbers unique and jarring.

Jain et al stated [103] "*Numbers have a language and give non-numerical perceptions*". The authors also went on to say there was no apparent reason for this kind of behaviour.

Baseline VAS Score	Count cases divisible by 5	Count cases divisible by 10	Twelve-mth VAS Score	Count cases divisible by 5	Count cases divisible by 10
0	3	3	0	0	0
5	0		5	0	
10	8	8	10	1	1
15	2		15	1	
20	15	15	20	4	4
25	14		25	2	
30	31	31	30	8	8
35	15		35	9	
40	60	60	40	14	14
45	19		45	5	
50	110	110	50	46	46
55	18		55	12	
60	70	70	60	47	47
65	32		65	24	
70	78	78	70	75	75
75	67		75	66	
80	68	68	80	108	108
85	10		85	55	
90	17	17	90	90	90
95	6		95	36	
100	9	9	100	24	24
Cases	652	469	**Cases**	627	417
Total Cases	819		**Total cases**	860	
Percentage	**79.61%**	**57.26%**	**Percentage**	**72.91%**	**48.49%**

Table 7 Visual Analogue Scores Baseline and 12-months grouped into divisible by 5 and 10

Baseline PL Score	Count cases divisible by 5	divisible by 10	Twelve-mth PL Score	Count cases divisible by 5	Count cases divisible by 10
0	2	2	0	1	1
5	0		5	0	
10	1	1	10	0	0
15	0		15	0	
20	24	24	20	1	1
25	0		25	1	
30	0	0	30	0	0
35	0		35	0	
40	28	28	40	17	17
45	2		45	0	
50	2	2	50	0	0
55	2		55	4	
60	0	0	60	0	0
65	0		65	2	
70	1	1	70	4	4
75	1		75	3	
80	28	28	80	81	81
85	0		85	1	
90	0	0	90	0	0
95	0		95	0	
100	1	1	100	4	4
Cases	92	87	Cases	119	108
Total Cases	815		Total Cases	855	
Percentage	11.29%	10.67%	Percentage	13.92%	12.63%

Table 8 Physical Limitation Baseline and 12-months grouped into divisible by 5 and 10

Table 9 is baseline and twelve-month descriptive statistics for mean, standard deviation and IQR for each of the SAQ domains and the VAS score. The numerator is shown for each domain. Both the VAS and individual SAQ domain scores range was between zero and one hundred.

With the exception of QL, all IQR's were either the same or lower at twelve-months than at baseline. The QL IQR increased from 25 to 33 at twelve-months indicating a slightly greater spread of scores.

Domain	Descriptor	n	Baseline	n	12 Months	% Change
VAS	Mean	819	57.79	860	74.33	28.62%
	Std Dev		19.96		16.67	
	IQR		30		22	
PL	Mean	815	48.06	855	63.83	32.81%
	Std Dev		20.14		18.05	
	IQR		29		29	
AS	Mean	787	46.25	770	80.42	73.88%
	Std Dev		31.58		26.66	
	IQR		50		50	
AF	Mean	782	54.42	833	88.47	62.57%
	Std Dev		27.74		18.92	
	IQR		50		20	
TS	Mean	801	81.78	828	87.47	6.96%
	Std Dev		14.10		13.72	
	IQR		18		12	
QL	Mean	809	38.65	832	71.91	86.05%
	Std Dev		20.46		21.84	
	IQR		25		33	

Table 9 Baseline and 12-month comparison between VAS and SAQ domains

n = numerator, PL = physical limitation, AS = angina stability, AF = angina frequency, TS = treatment satisfaction, QL = quality of life, IQR = interquartile range

5.7 Baseline Histograms

Baseline histograms were created for VAS (Figure 9), PL (Figure 10), AS (Figure 11), AF (Figure 12), TS (Figure 13) and QL (Figure 14).

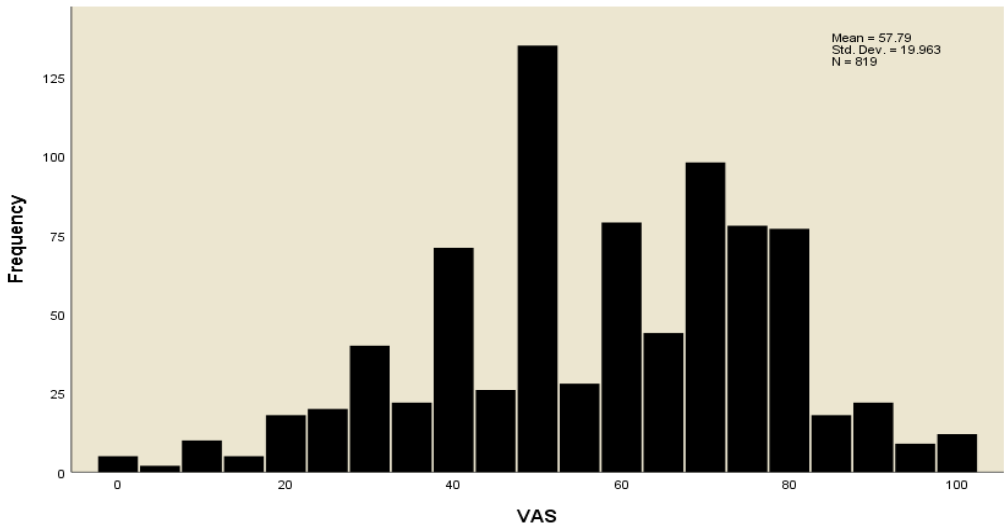


Figure 9 Baseline histogram: Visual Analogue Scale

Numerator = 819, Mean = 57.79, Standard deviation = 19.96, Interquartile range = 30

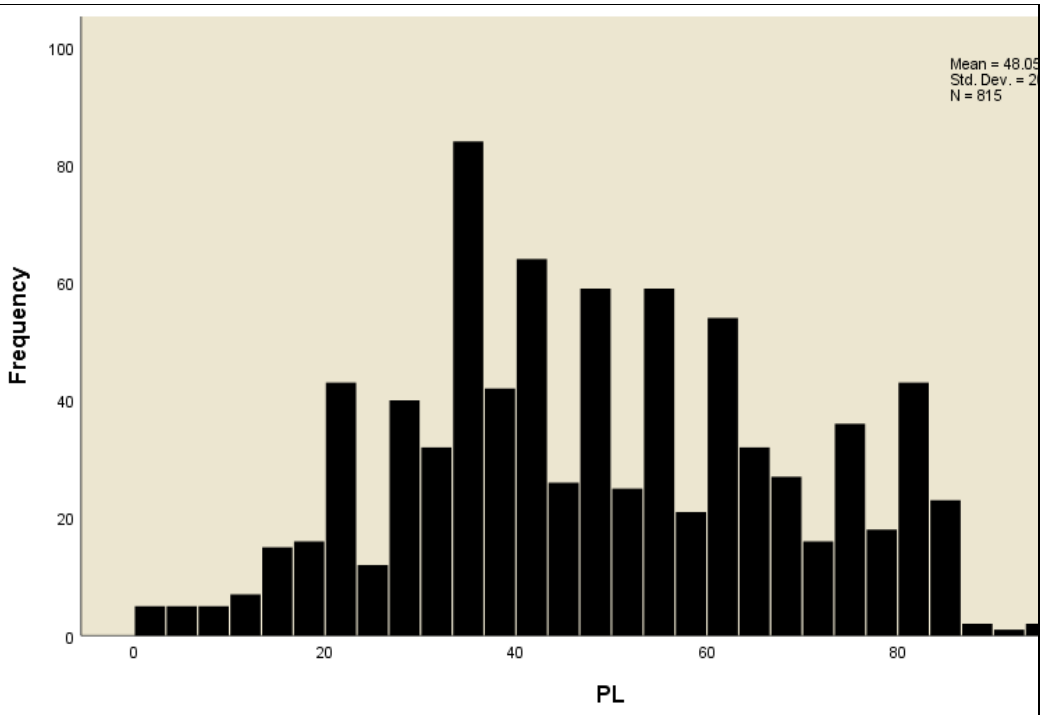


Figure 10 Baseline histogram: Physical Limitation

Numerator = 815, Mean = 48.06, Standard deviation = 20.14, Interquartile range = 29

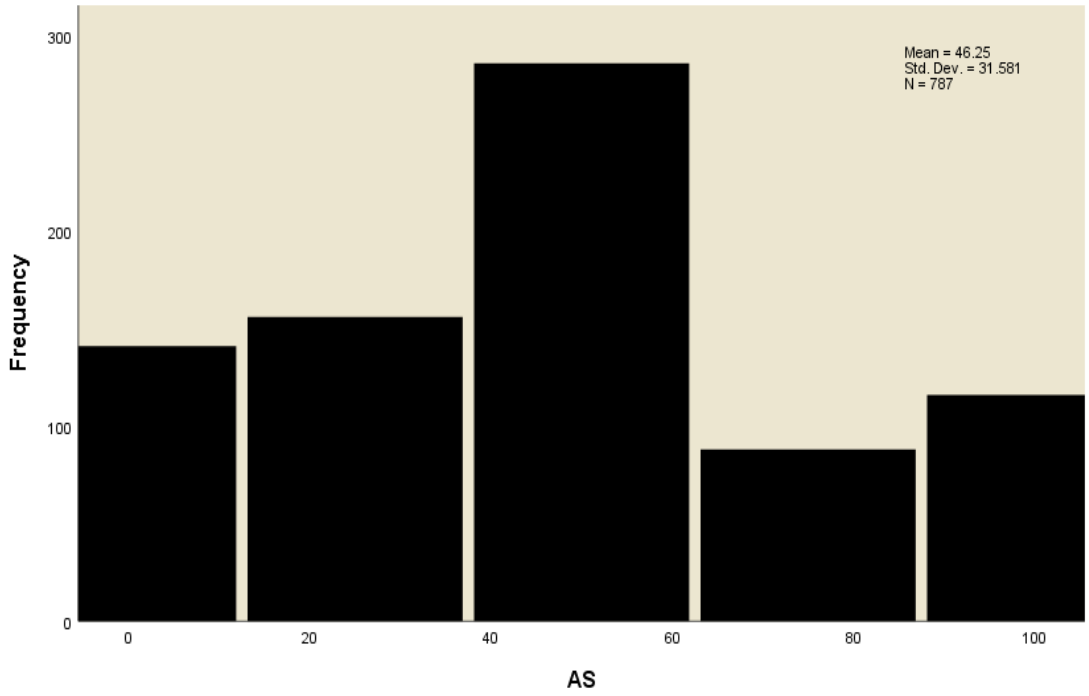


Figure 11 Baseline Histogram: Angina Severity

Numerator = 787, Mean = 46.25, Standard deviation = 31.58, Interquartile range = 50

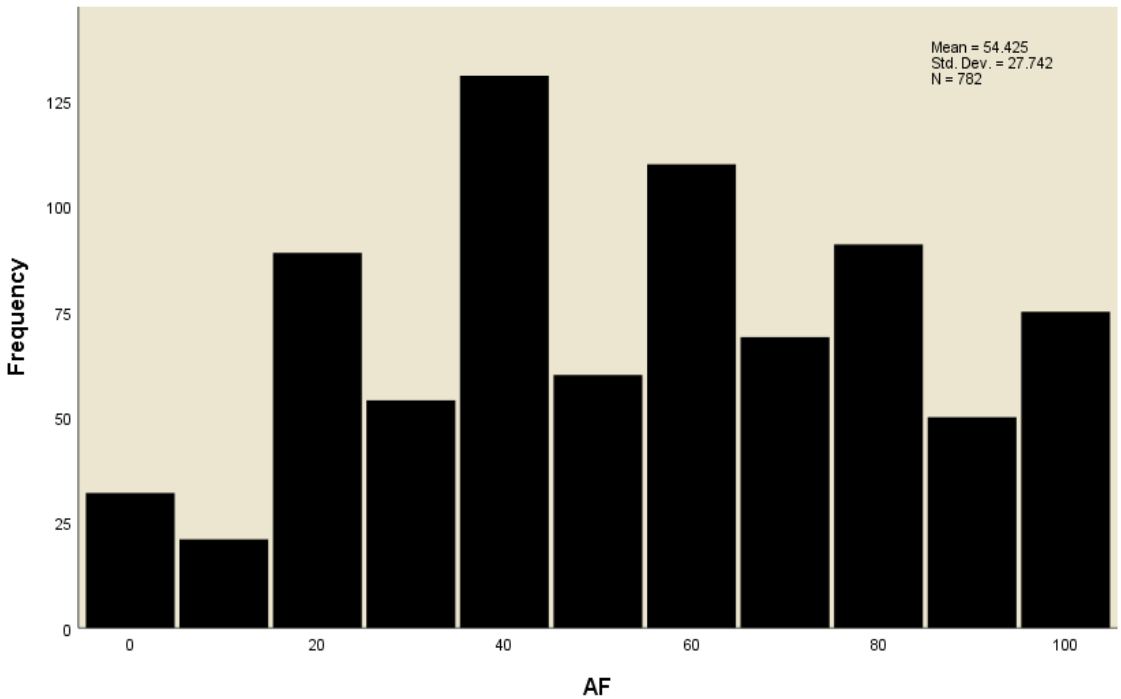


Figure 12 Baseline Histogram: Angina Frequency

Numerator = 782, Mean = 54.42, Standard deviation = 27.74, Interquartile range = 50

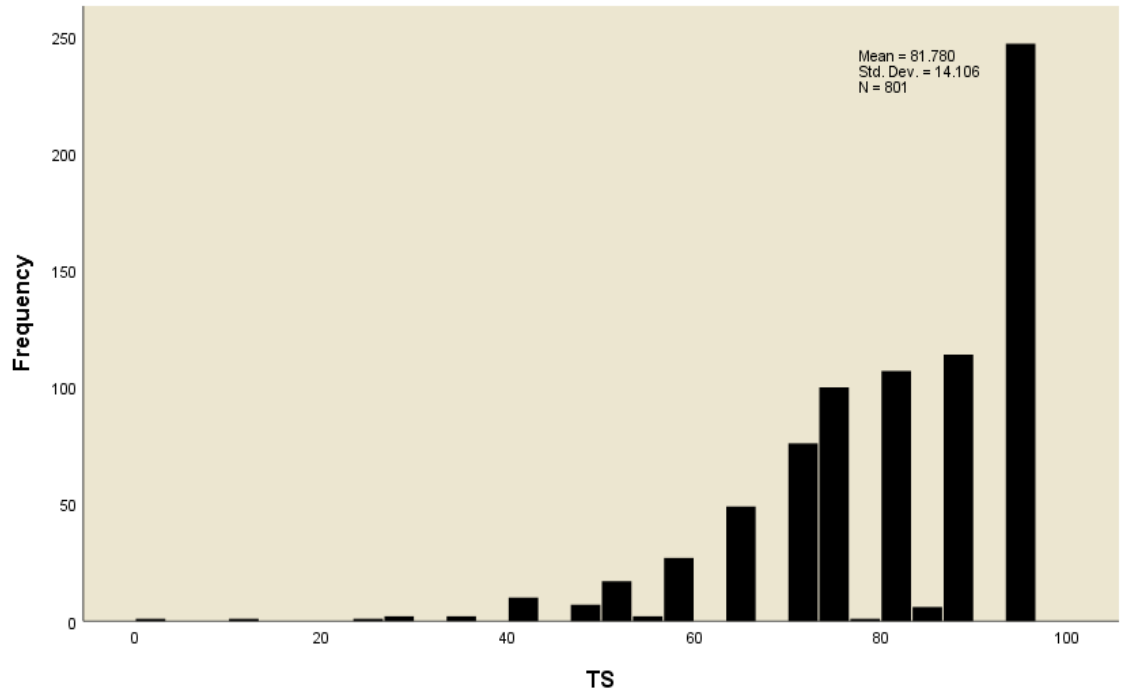


Figure 13 Baseline Histogram: Treatment Satisfaction

Numerator = 801, Mean = 81.78, Standard deviation = 14.10, Interquartile range = 18

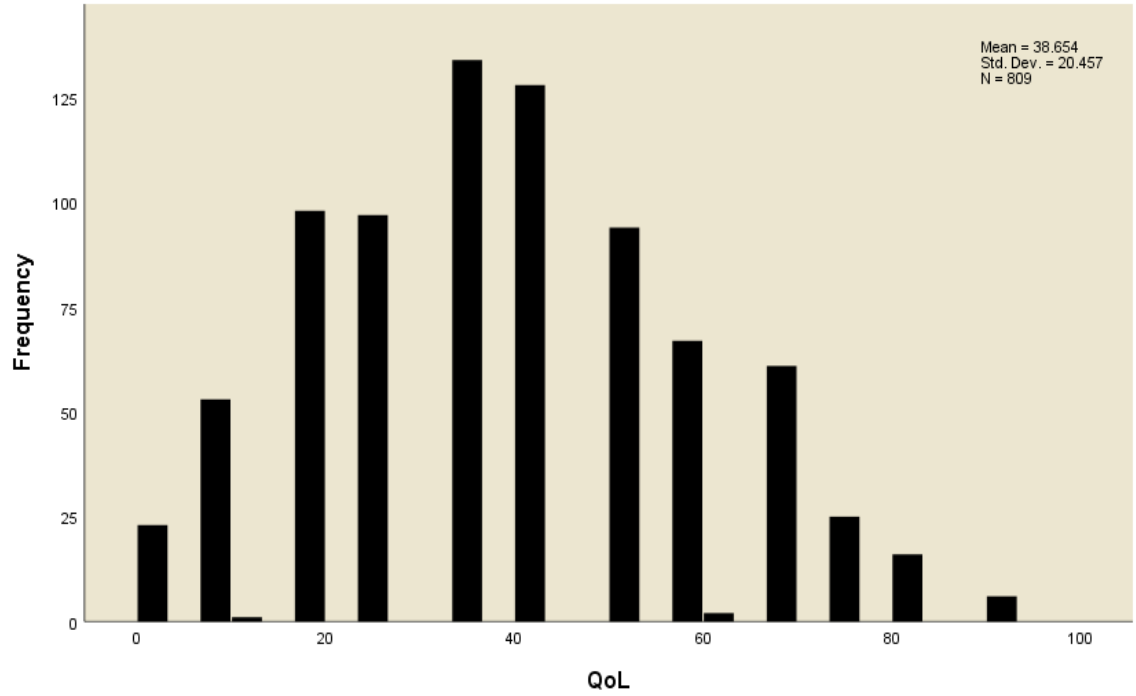


Figure 14 Baseline Histogram: Quality of Life

Numerator = 809, Mean = 38.65, Standard deviation = 20.46, Interquartile range = 25

5.8 Twelve-Month Histograms

Twelve-month histograms were created for VAS (Figure 15), PL (Figure 16), AS (Figure 17), AF (Figure 18), TS (Figure 19) and QL (Figure 20).

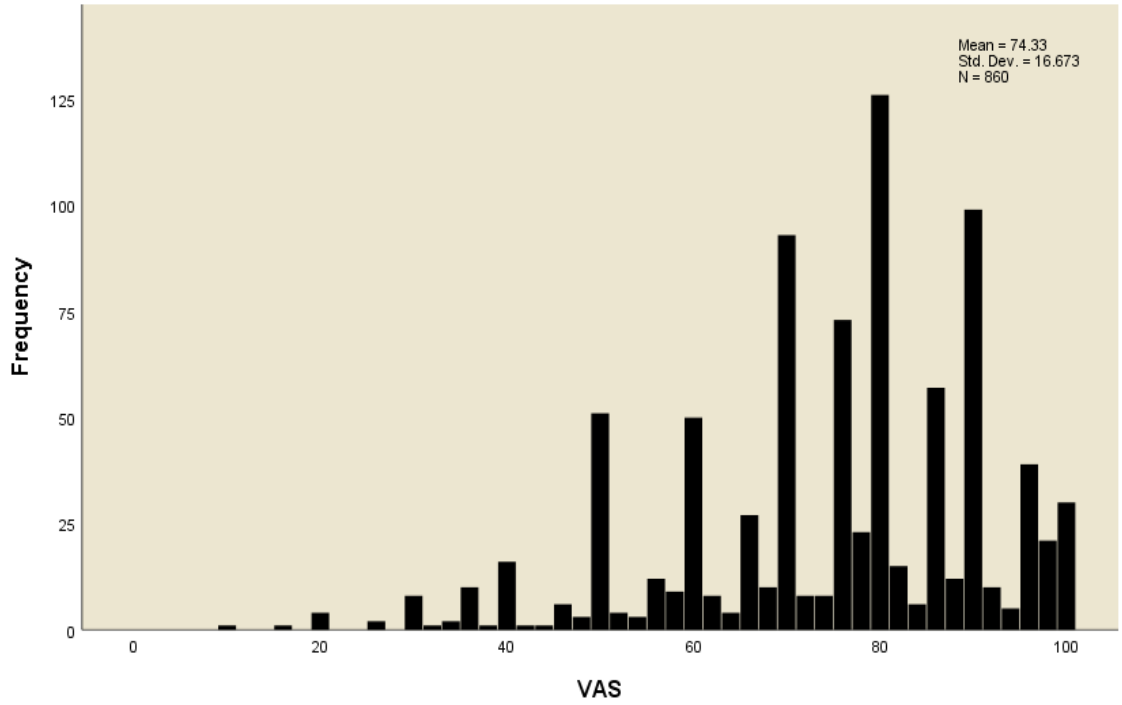


Figure 15 Twelve-months Histogram: Visual Analogue Scale

Numerator = 860, Mean = 74.33, standard deviation = 16.67, interquartile range = 22

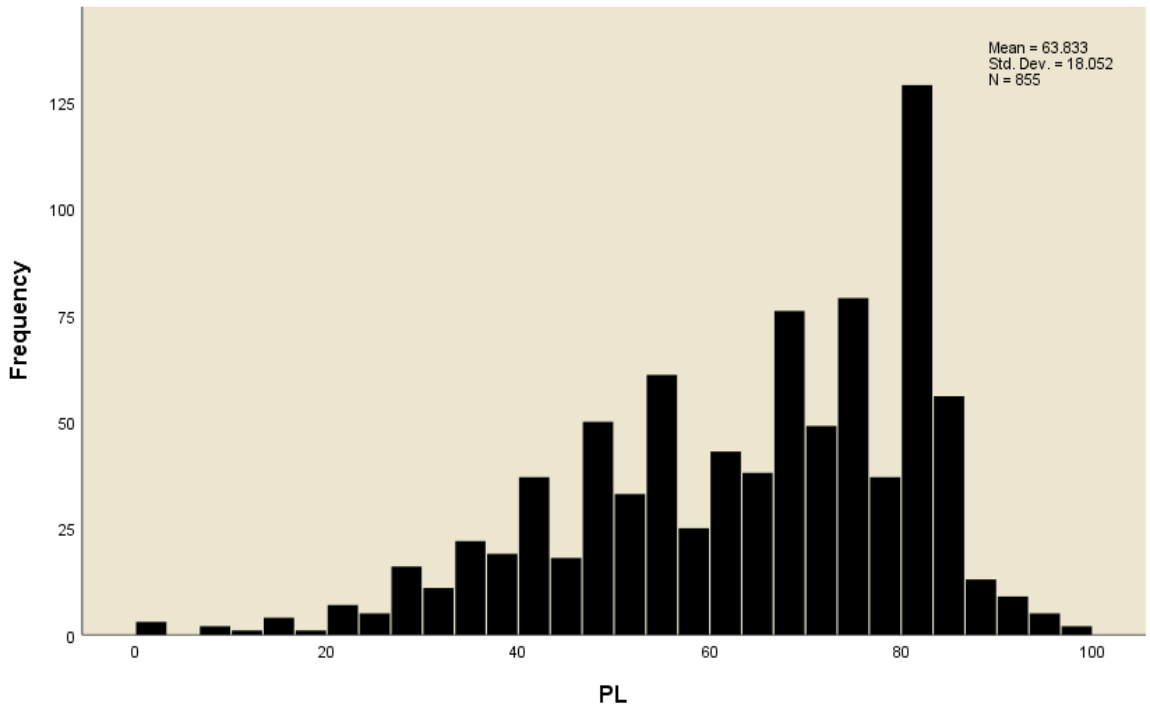


Figure 16 Twelve-months Histogram: Physical Limitation

Numerator = 855, Mean = 63.83, Standard deviation = 18.05, Interquartile range = 29

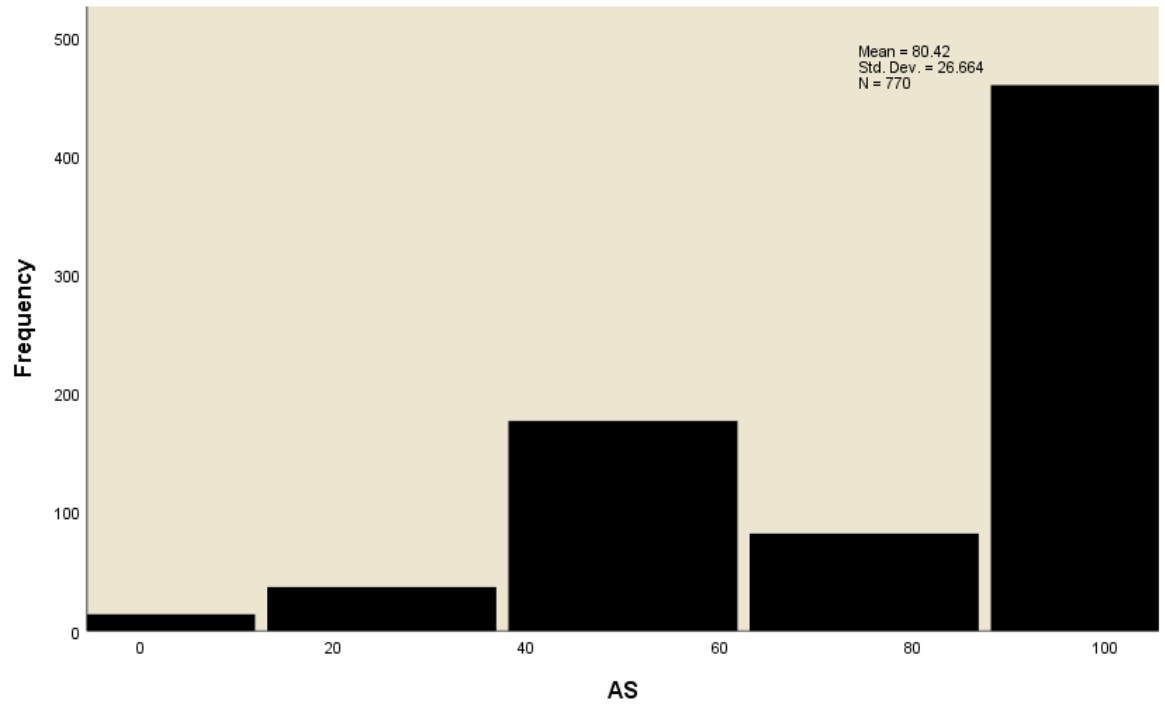


Figure 17 Twelve-months Histogram: Angina Stability

Numerator = 770, Mean = 80.42, Standard deviation = 26.66, Interquartile range = 50

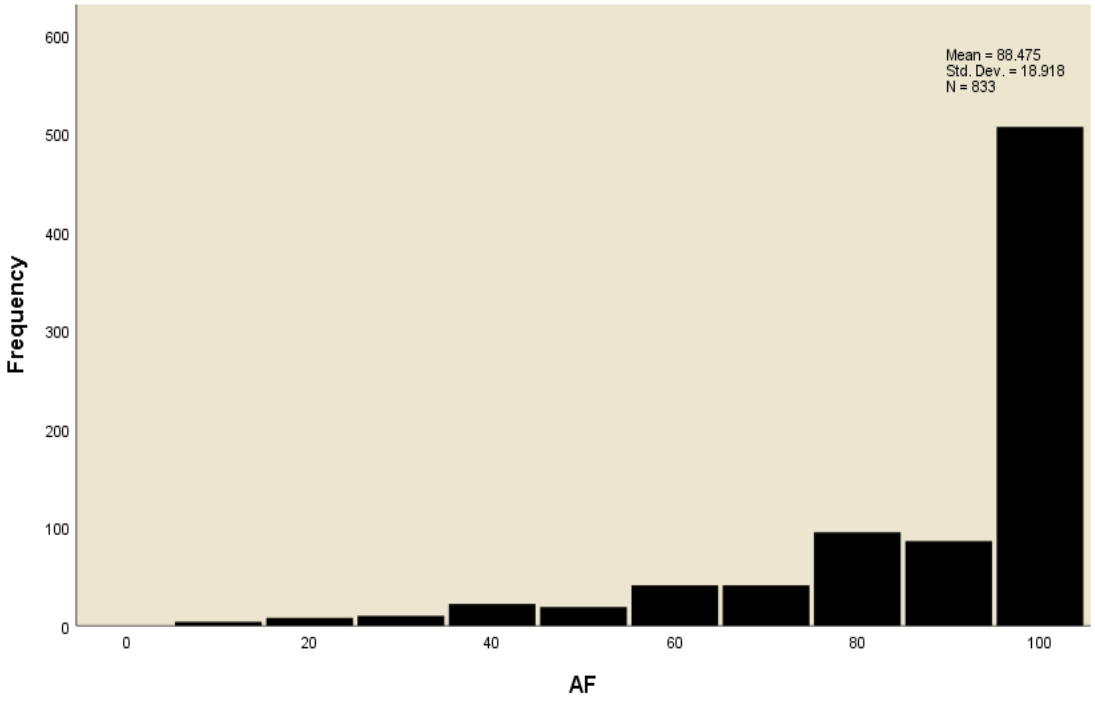


Figure 18 Twelve-months Histogram: Angina Frequency

Numerator = 833, Mean = 88.47, Standard deviation = 18.92, Interquartile range = 20

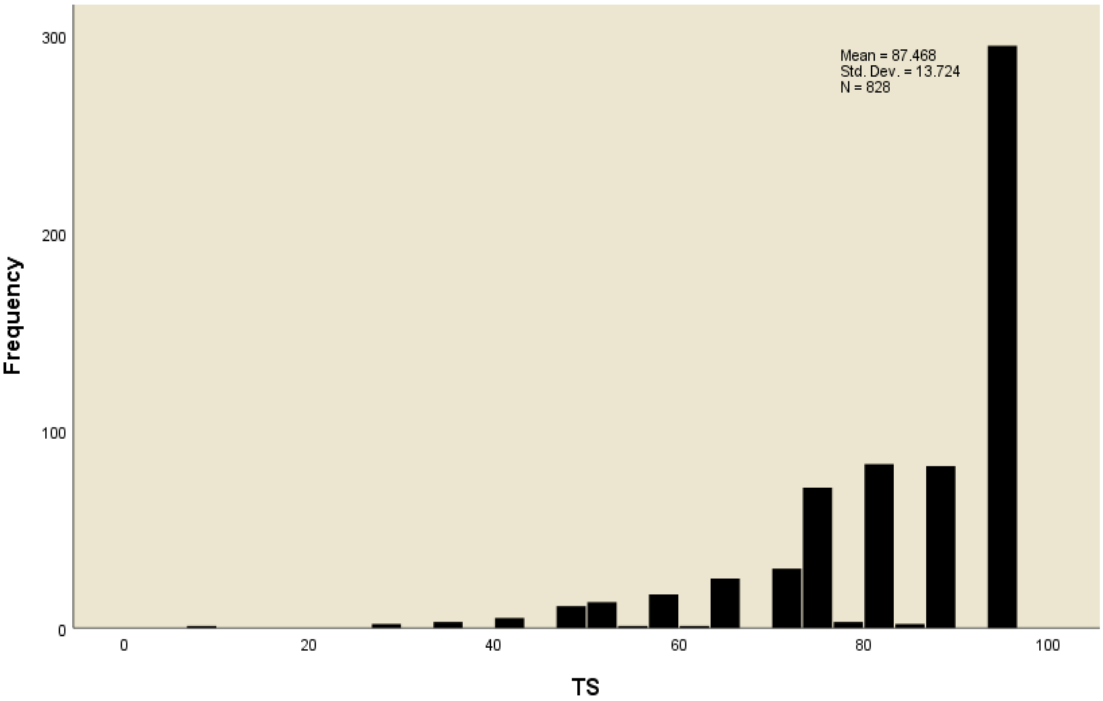


Figure 19 Twelve-months Histogram: Treatment Satisfaction

Numerator = 828, Mean = 87.47, Standard deviation = 13.72, Interquartile range = 33

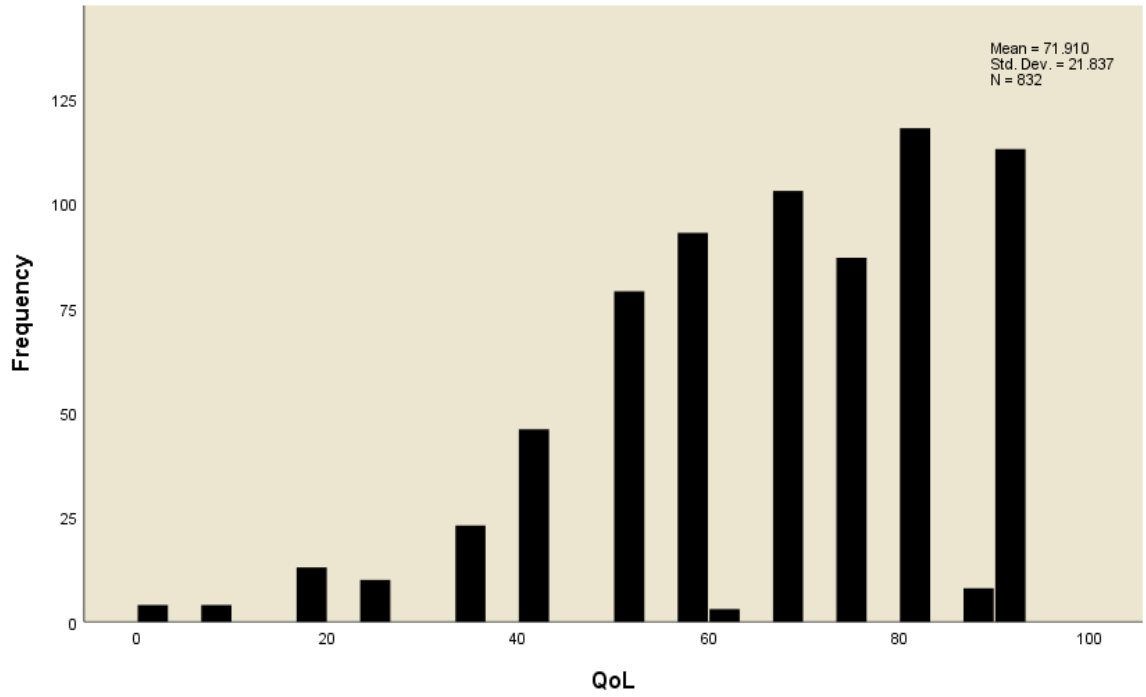


Figure 20 Twelve-months Histogram: Quality of Life
Numerator = 832, Mean = 71.91, Standard deviation = 21.84, Interquartile range = 33

5.9 Magnitude of Change Baseline to Twelve Months

Histograms were also created to display the magnitude of change by subtracting the baseline from twelve-months data.

5.9.1 Magnitude of Change - VAS

Figure 21 relates to the Visual Analogue Scale. The numerator was 722 and the mean is 16.95, indicating a positive change over the year. However, the standard deviation is 21.49.

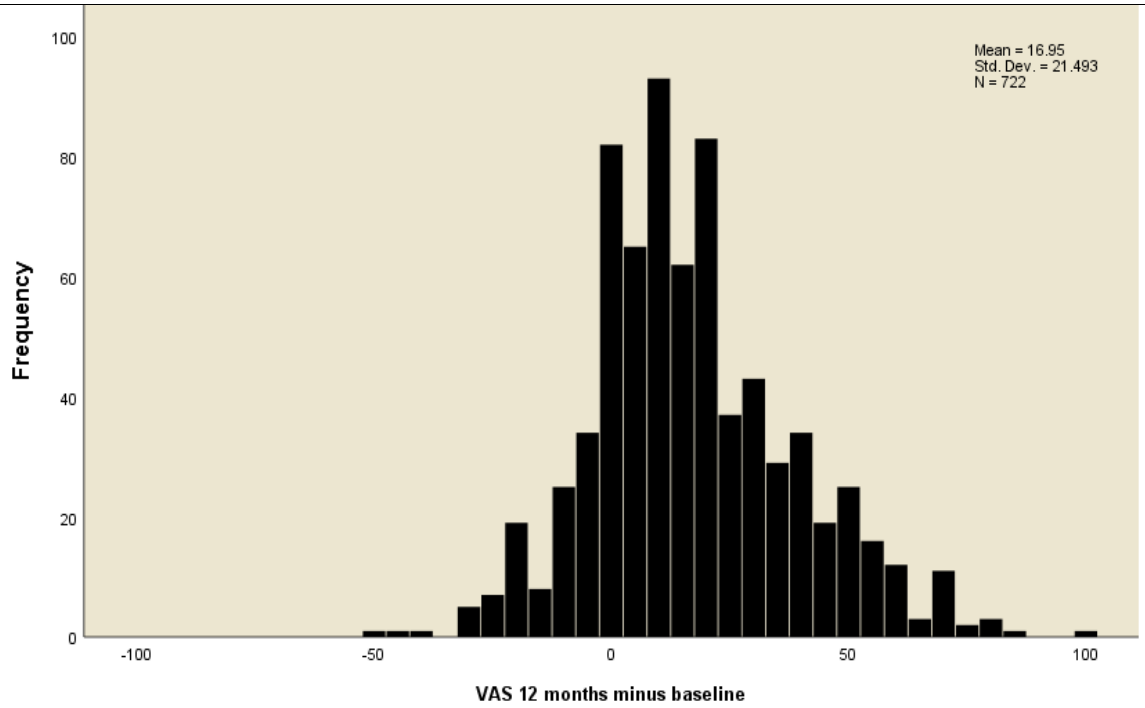


Figure 21 Magnitude of Change Over 12 Months: Visual Analogue Scale
Mean = 16.95, Standard Deviation = 21.49, Numerator = 722

5.9.2 Magnitude of Change – Physical Limitation

Figure 22 relates to Physical Limitation. The numerator here is 877 and the mean is 16, indicating a positive change over the year. There is a wide standard deviation of 22.1.

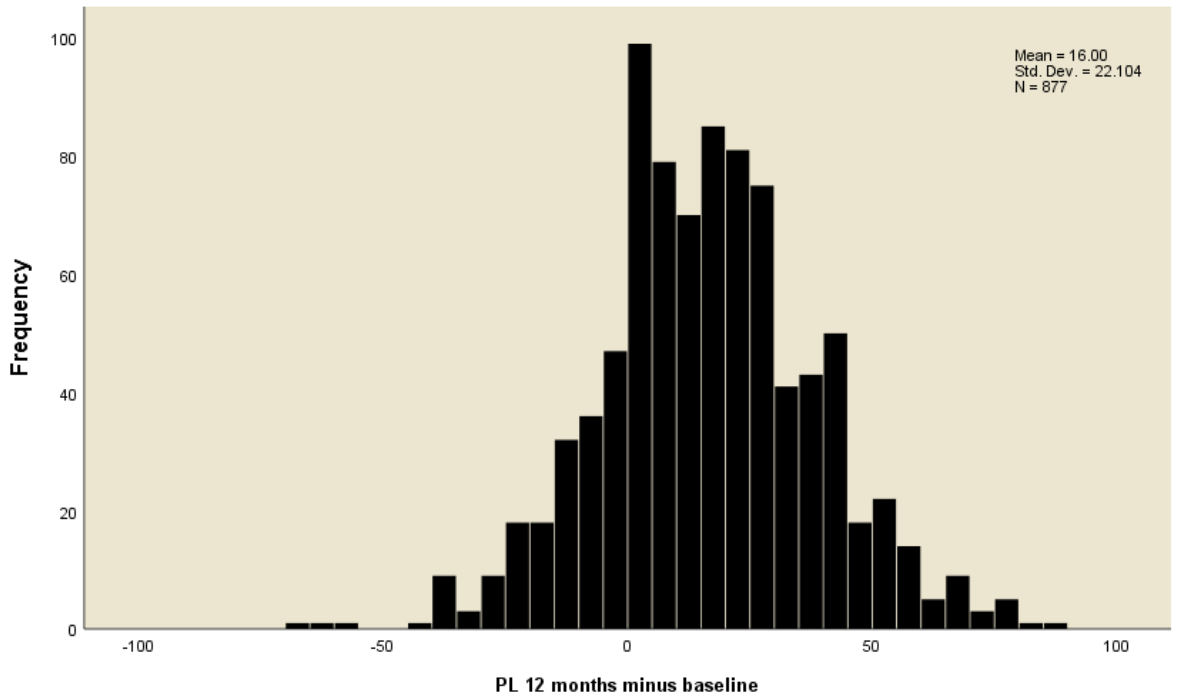


Figure 22 Magnitude of Change Over 12 Months: Physical Limitation
Mean = 16, Standard Deviation = 22.104, Numerator = 877

5.9.3 Magnitude of Change – Angina Stability

Figure 23 relates to angina stability. The numerator is 764 and the mean is 35.5 indicating more stable angina over the year. However, the standard deviation was wide at 37.99. There were cases at both extremes of minus 100 and plus 100.

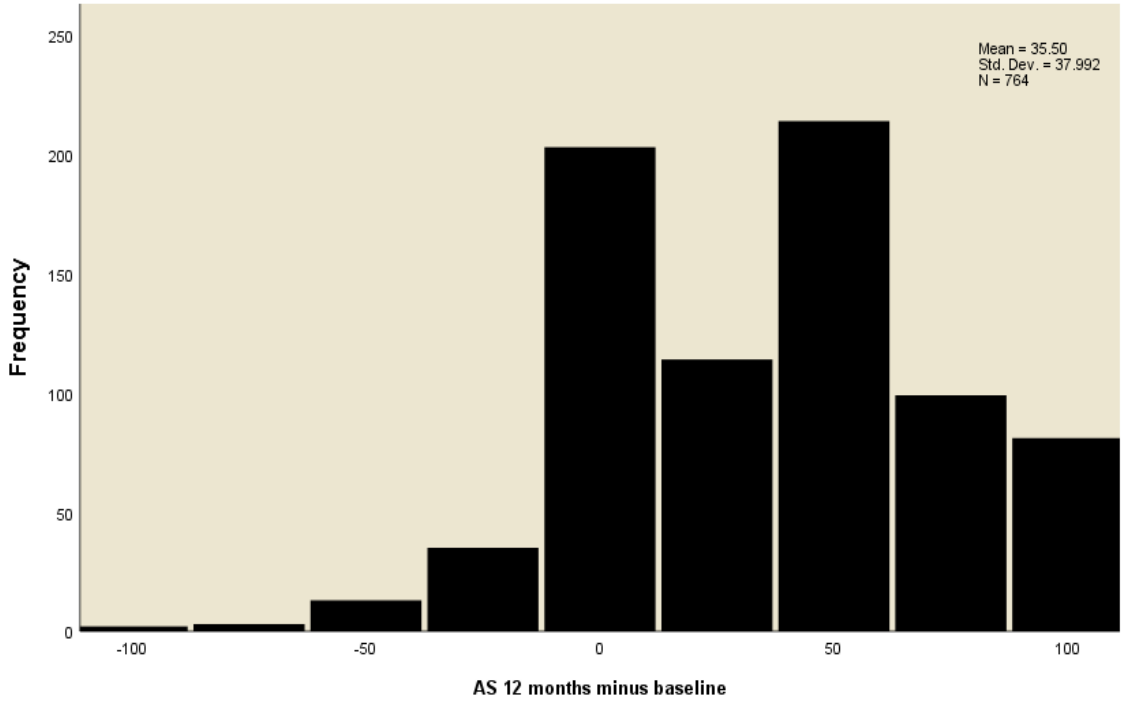


Figure 23 Magnitude of Change Over 12 Months: Angina Stability
Mean = 35.50, Standard Deviation = 37.992, Numerator = 764

5.9.4 Magnitude of Change – Angina Frequency

Figure 24 relates to angina frequency. The numerator is 812 and the mean is 33.63, an improvement over the twelve-months. Again, there is a wide standard deviation of 29.

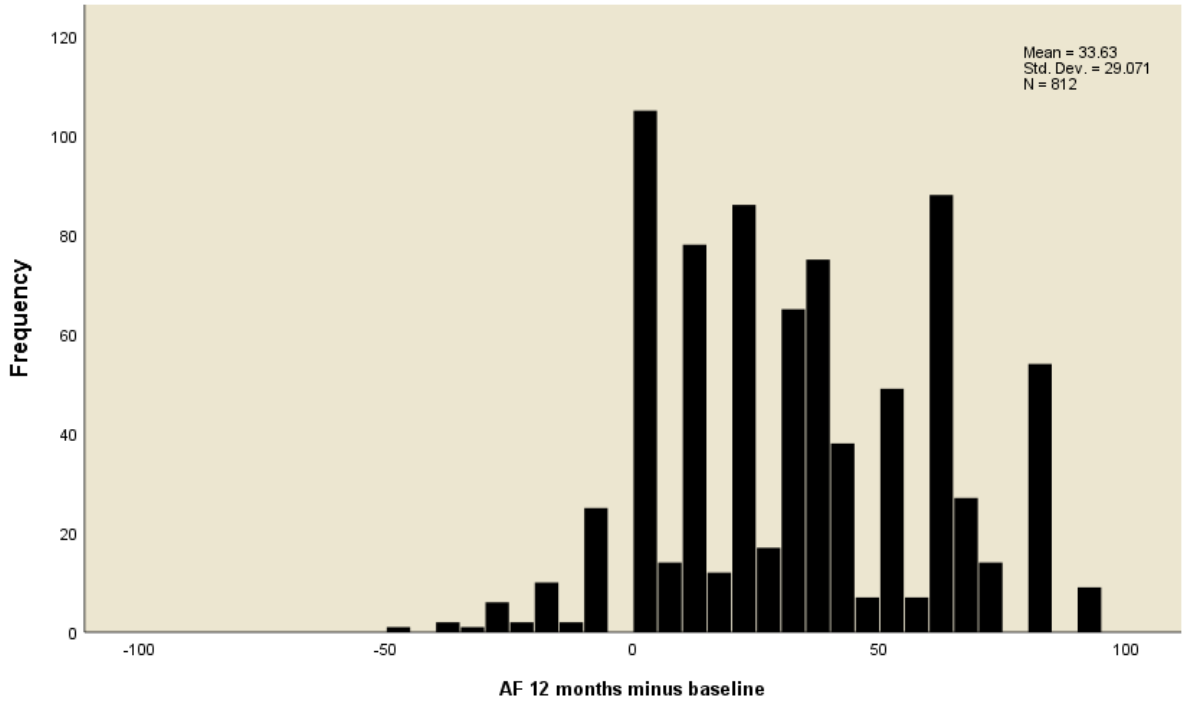


Figure 24 Magnitude of Change Over 12 Months: Angina Frequency
Mean = 33.63, Standard Deviation = 29.071, Numerator = 812

5.9.5 Magnitude of Change – Treatment Satisfaction

Figure 25 relates to treatment satisfaction. The numerator is 834 and mean is

5.56. Although there is an improvement in treatment satisfaction it is modest.

There is a standard deviation of 16, narrower than the other cases.

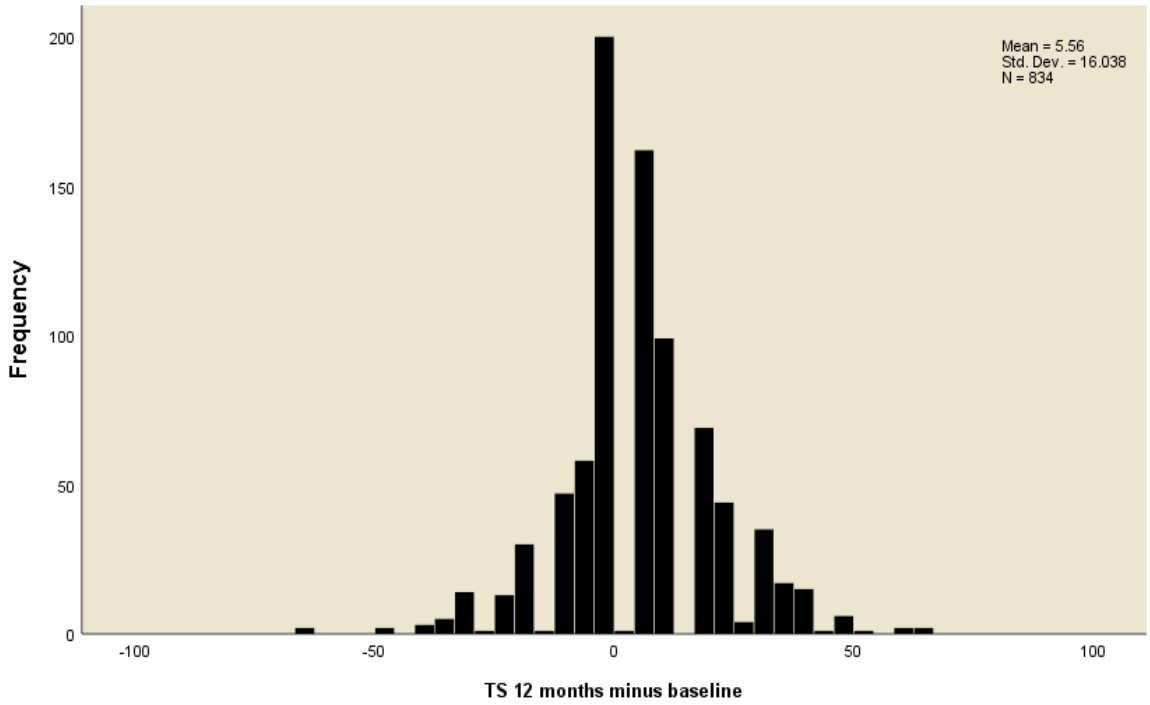


Figure 25 Magnitude of Change Over 12 Months: Treatment Satisfaction
Mean = 5.56, Standard Deviation = 16.036, Numerator = 834

5.9.6 Magnitude of Change –Quality of Life

Figure 26 relates to quality of life. The numerator is 846 and the mean is 34. The mean is only slightly lower than for angina stability, and the standard deviation is 24.8.

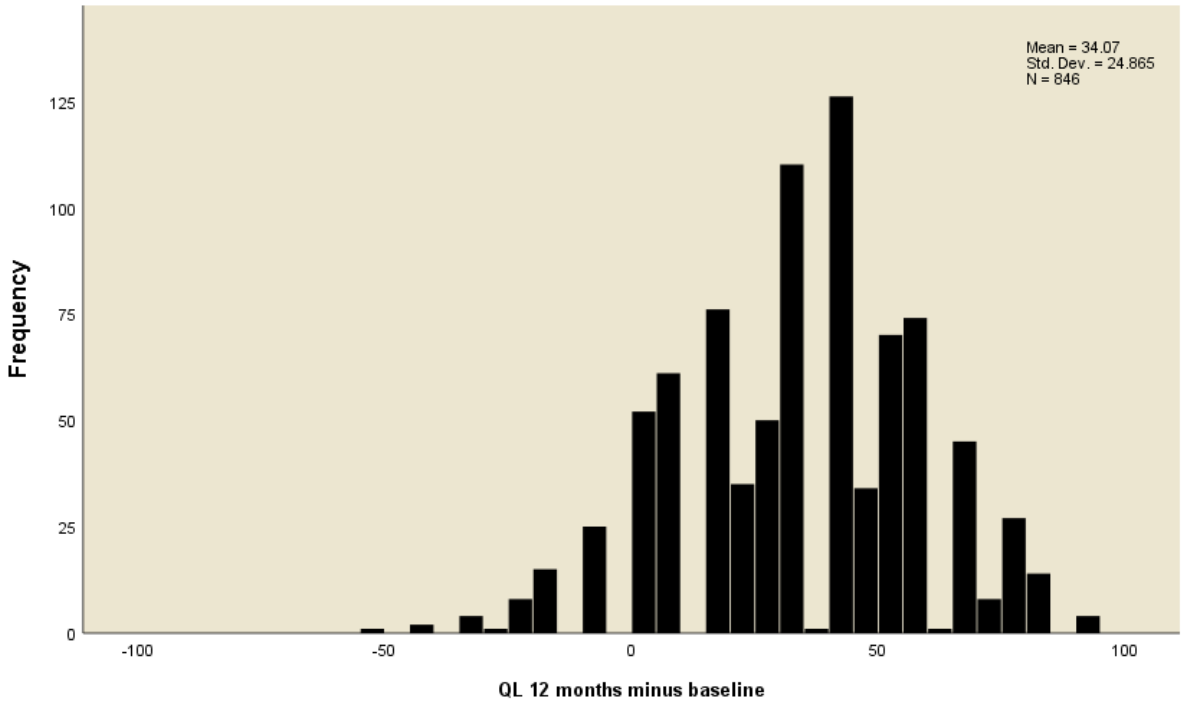


Figure 26 Magnitude of Change Over 12 Months: Quality of Life
Mean = 34.07, Standard Deviation = 224.86, Numerator = 846

5.10 Bland and Altman Plots

Bland and Altman plots were created to compare the differences between the VAS and each domain score of the SAQ with the averages of the two reported scores. Bland and Altman plots do not say if agreement is sufficient or if one or the other method of scoring should be used. It simply quantifies the bias and range of agreement between the measures [(104)].

To create these plots in SPSS I first needed to create two additional variables for each domain of the SAQ. One was the difference between the VAS and each domain of the SAQ and the other was the average of the two reported scores.

To calculate the upper confidence interval (CI), I used the formula (Std Dev x 1.96) + mean. The lower confidence interval was calculated using the formula: Mean - (Std Dev x 1.96).

I then created the plots in SPSS which start as a scatter plot. The 'Y' axis was the difference between the two variables and the 'X' axis the mean of the two variables. In the chart editor I then added three lines using the 'Y-axis reference line' tool. I then entered the value for the mean calculated in Excel. I repeated this process for the upper and lower CI limit lines.

5.11 Scatter Plots

A scatter plot will show the relationship between two variables and how strongly the two variables are related. However, a high correlation does not imply there is good agreement between the two variables. Correlation is concerned with the strength of a relationship, not the agreement between them [(104)].

For clarity, the various Bland and Altman and scatter plots will be described as they relate separately for each domain of the SAQ. The descriptions will include baseline and twelve-months. Both the magnitude of change and the extent of agreement will be described.

5.12 VAS and Physical Limitation

5.12.1 Bland Altman

Bland and Altman proposed a method of measuring the agreement between two measurements or raters by measuring the mean differences [(104)]. This differs from correlation, which measures how strongly two measures are related, not their differences. High levels of correlation do not imply good agreement. On a Bland and Altman plot, the Y axis represents the difference between two measures and the X axis represents the mean average of the two measurements. A Bland and Altman plot does not state if the agreement is suitable to use one or

the other measures. The plots also contain a central line indicating the mean and an upper and lower 95% confidence interval line.

Figure 27 is the Bland Altman plot displaying baseline data. The mean is 9.82 indicating a bias towards VAS and the majority of data points fall between the upper and lower CI. However, the CI's are wide, between 52 and -33. There is a wide spread of data points on the 'x' axis from 0 to 90 which you would expect at

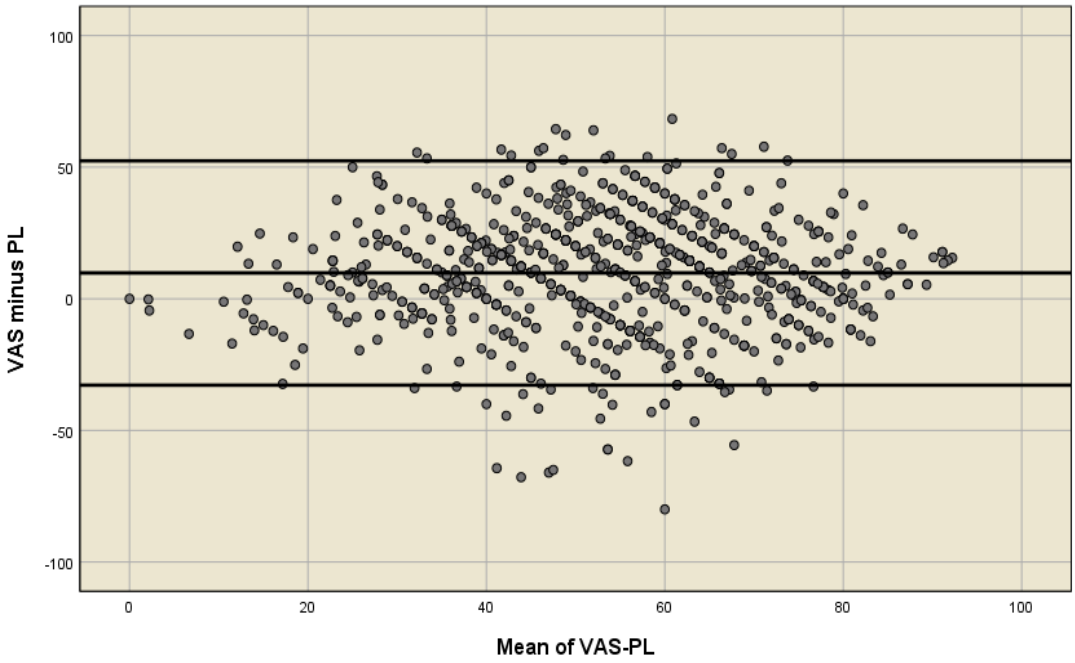


Figure 27 Bland Altman Plot Baseline: VAS and Physical Limitation
Mean = 9.82, Upper 95% CI is 52, Lower 95% CI is -33

baseline prior to revascularisation.

Figure 28 is the twelve-month Bland Altman plot. The mean is 10.46 so the bias favouring VAS has increased and again there are wide CI's at 46 to -25, although the CI's are not as wide as baseline. The majority of the data points are above 60 on the 'x' axis. This would be expected as twelve-months after revascularisation the physical limitations should not be worse than at baseline.

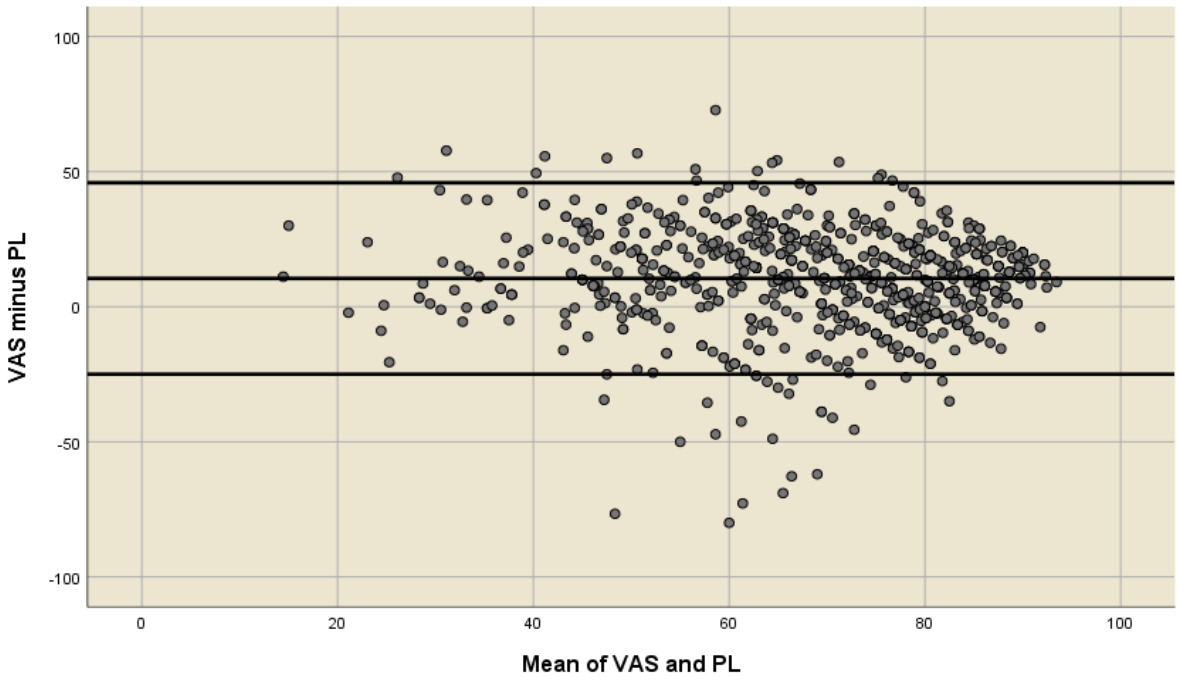


Figure 28 Bland Altman plot twelve-months: VAS and Physical Limitation
Mean = 10.46, Upper 95% CI = 46, Lower 95% CI = -25

5.12.2 Scatter Plots

Figure 29 is the scatter plot for VAS and PL at baseline. The data range of both axis is wide from 0 to 100. The correlation coefficient is 0.167, showing a weak positive correlation.

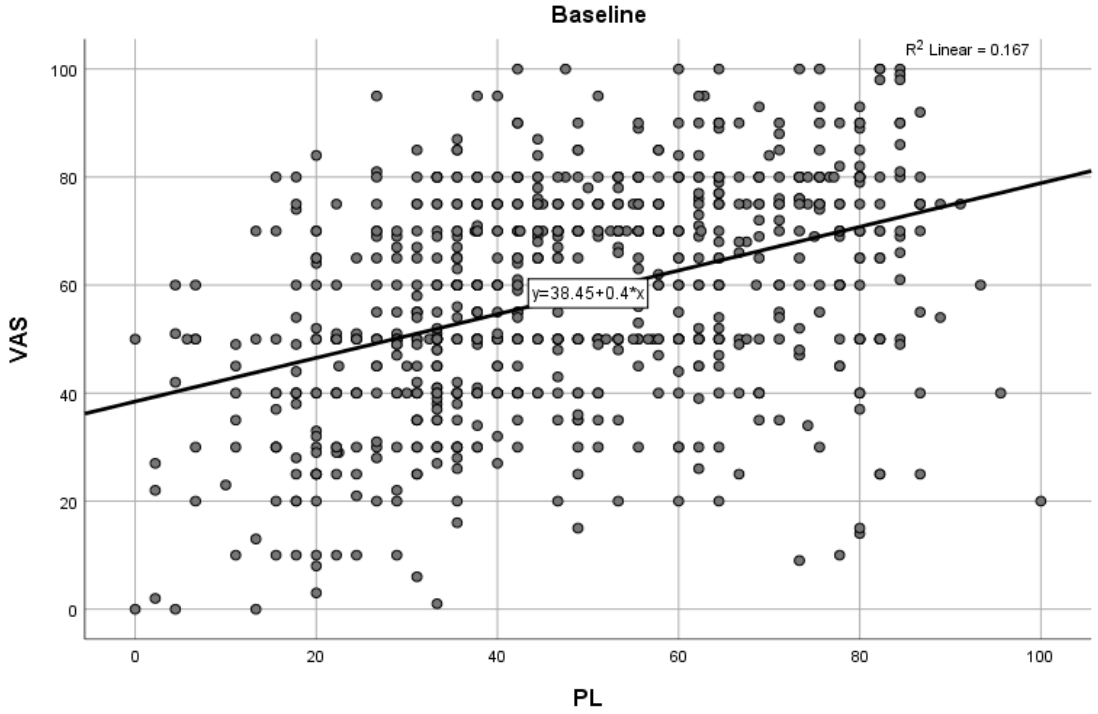


Figure 29 Scatter plot baseline: VAS and Physical Limitation
Correlation coefficient = 0.167

Figure 30 is the twelve-month data. This shows a slightly stronger correlation coefficient of 0.213. Data points are more common above 60.

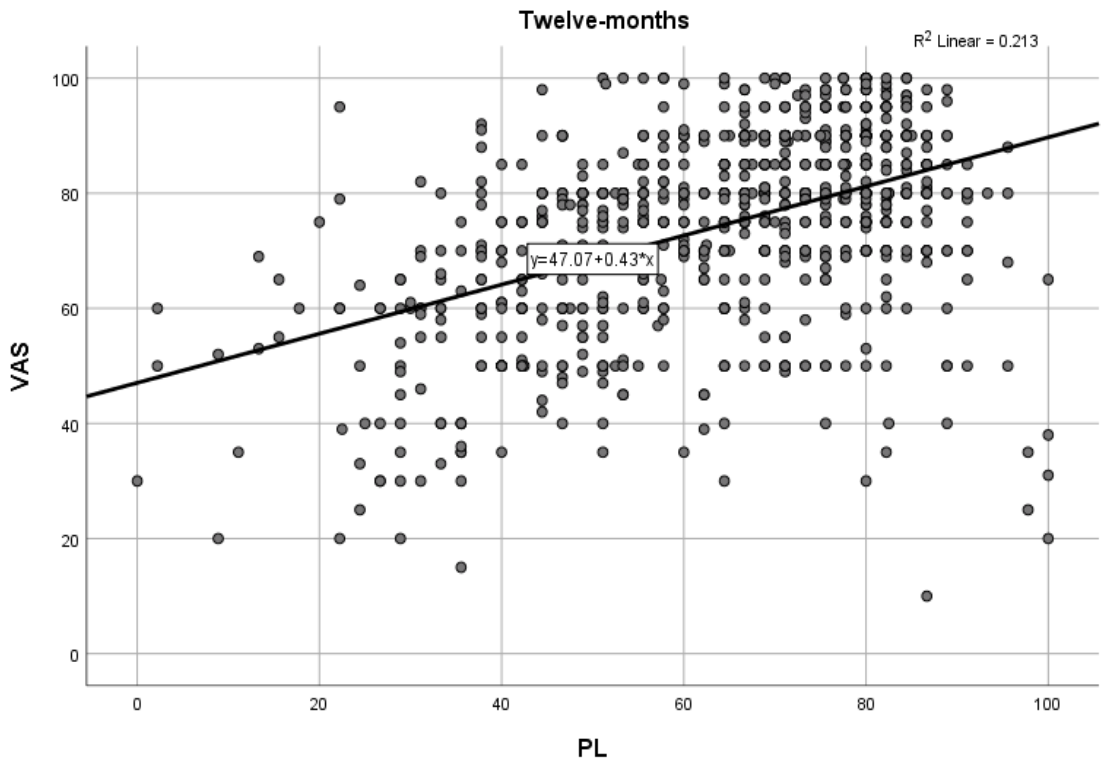


Figure 30 Scatter plot twelve-months: VAS and Physical Limitation
Correlation coefficient = 0.213

Figure 31 is a scatter plot displaying the magnitude of change over twelve-months. Although there is still a low correlation coefficient of 0.124, the majority of data points are clumped together centrally.

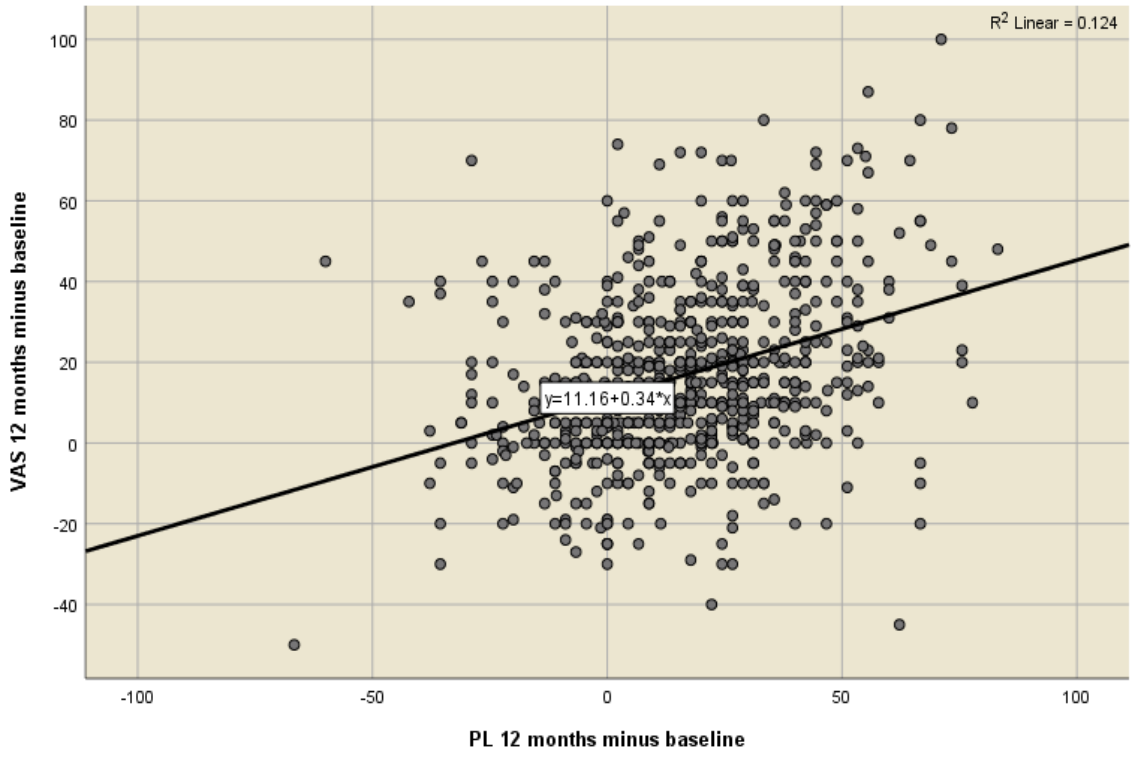


Figure 31 Change from baseline to 12 months: VAS and Physical Limitation
Correlation coefficient = 0.124

5.13 VAS and Angina Stability

5.13.1 Bland Altman

Figure 32 displays baseline data. The mean is 11.47 indicating a bias towards VAS but there are very wide 95% confidence intervals of between 72 and -49. There is a very wide range of data points on the mean axis.

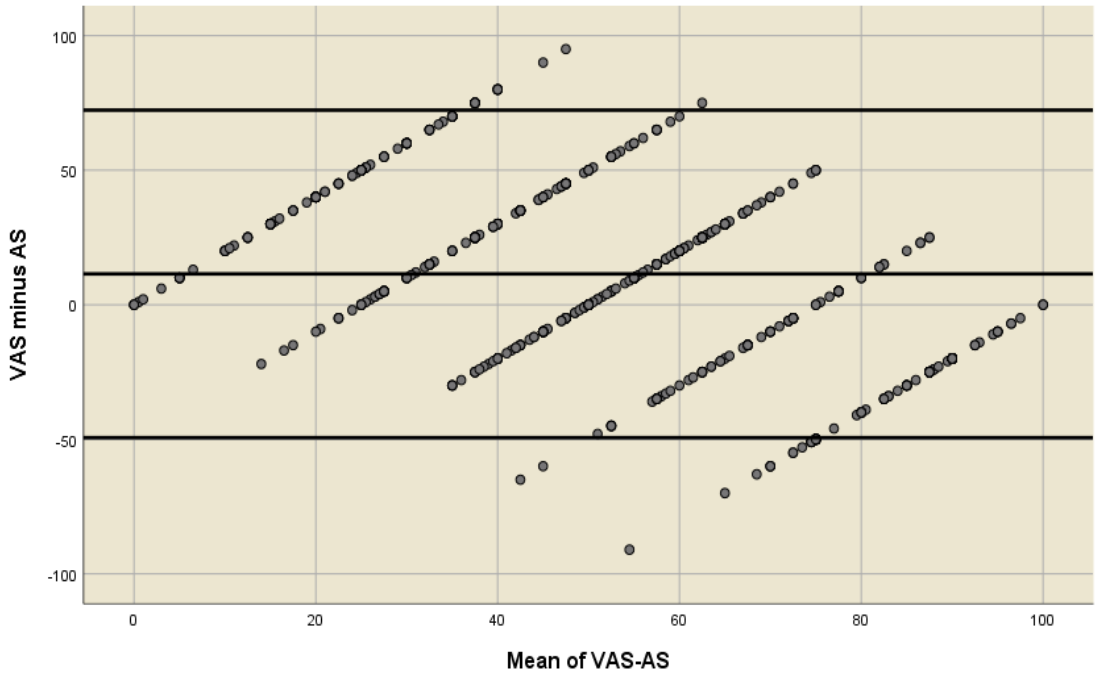


Figure 32 Bland Altman plot baseline: VAS and Angina Stability
Mean = 11.47, Upper 95% CI is 72, Lower 95% CI is -49

Figure 33 for twelve months shows a negative mean of -7.16, a bias in favour of AS which is a reversal of the baseline position and the 95% CI's are lower at 43 to -57.

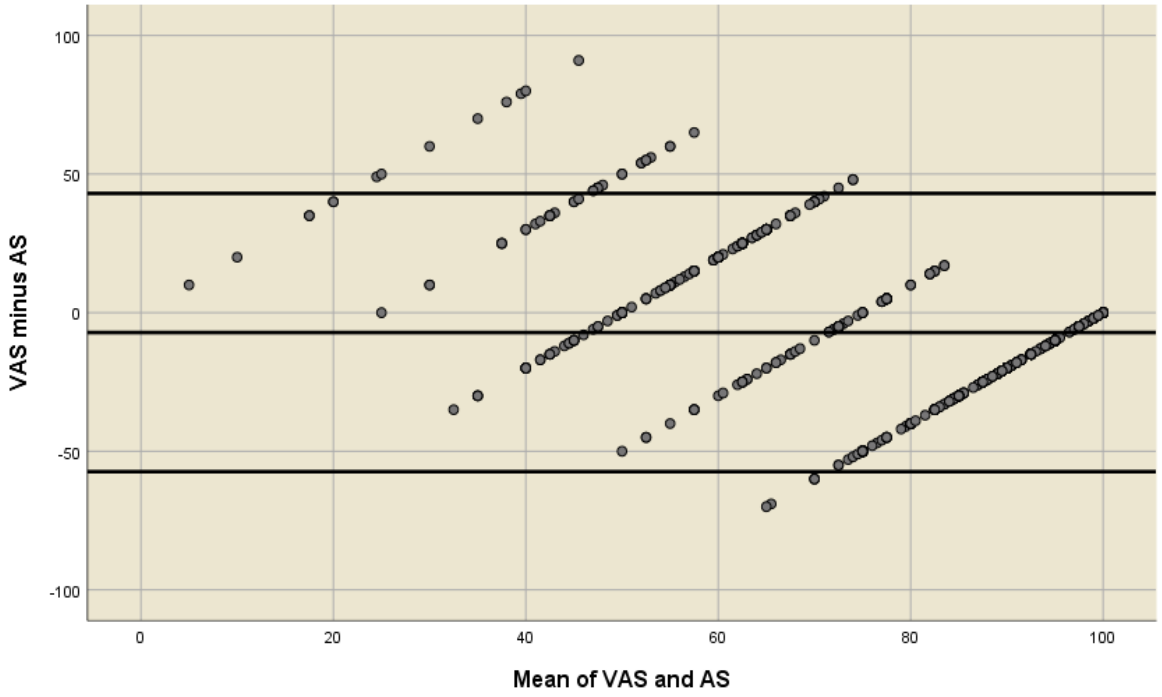


Figure 33 Bland Altman plot twelve-months: VAS and Angina Stability
Mean = -7.16, Upper 95% CI = 43, Lower 95% CI = -57

5.13.2 Scatter Plots

Figure 34 is the baseline scatter plot for VAS and AS. There is a weak correlation coefficient of 0.115.

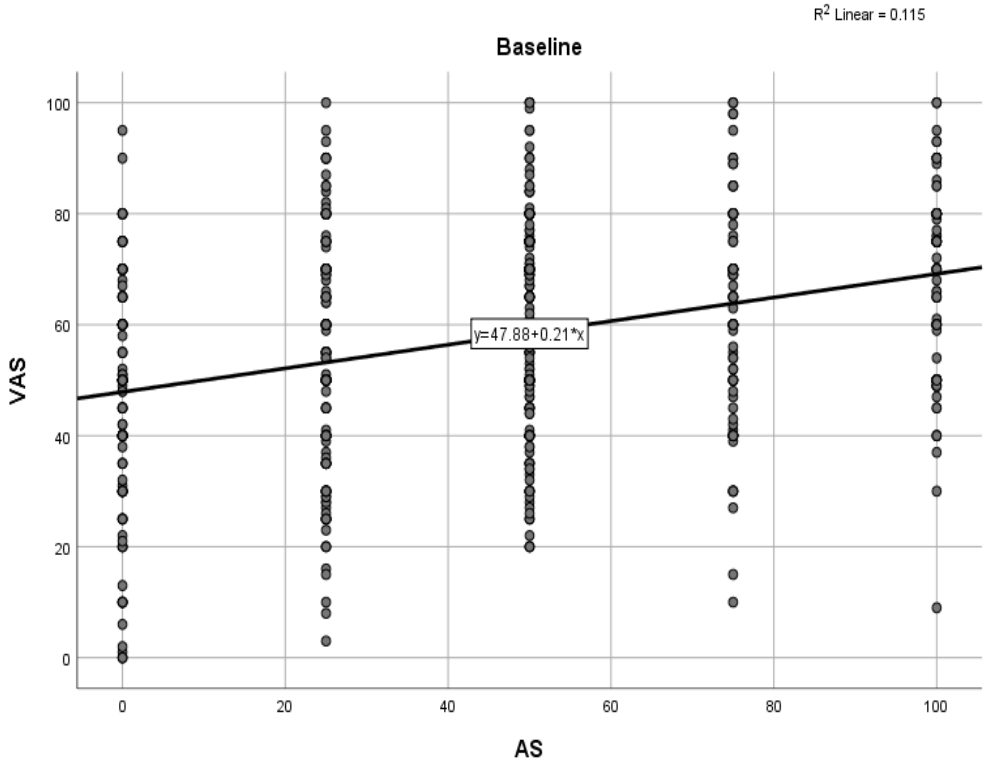


Figure 34 Scatter plot baseline: VAS and Angina Stability
Correlation coefficient = 0.115

Figure 35 relates to twelve-months and shows a similar weak correlation coefficient of 0.142.

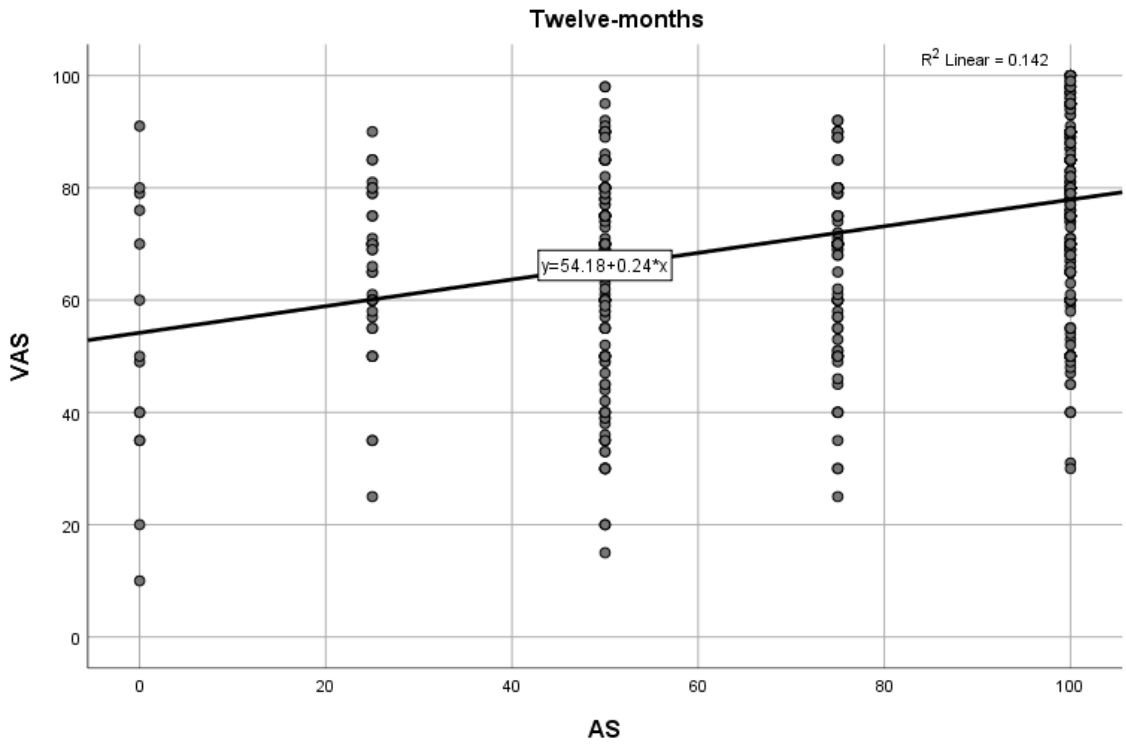


Figure 35 Scatter plot twelve-months: VAS and Angina Stability
Correlation coefficient = 0.142

Figure 36 is the magnitude of change over twelve-months. Again, there is a very weak correlation coefficient of 0.086 indicating almost no relationship between VAS and AS in terms of the magnitude of change.

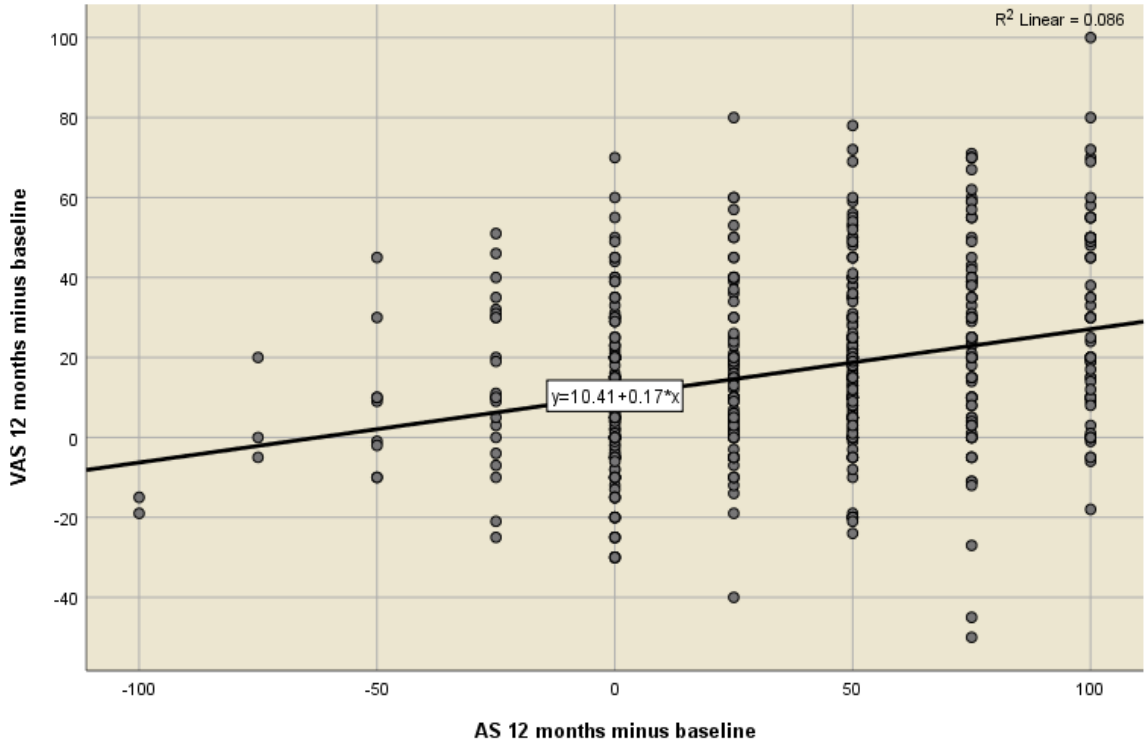


Figure 36 Change from baseline to 12 months: VAS and Angina Stability
Correlation coefficient = 0.086

5.14 VAS and Angina Frequency

5.14.1 Bland Altman

Figure 37 is the Bland Altman for baseline. The mean is 3.24 indicating a small bias in favour of VAS. There are very wide confidence intervals of 59 to -52 and the data points range from 0 to 100 on the 'x' axis.

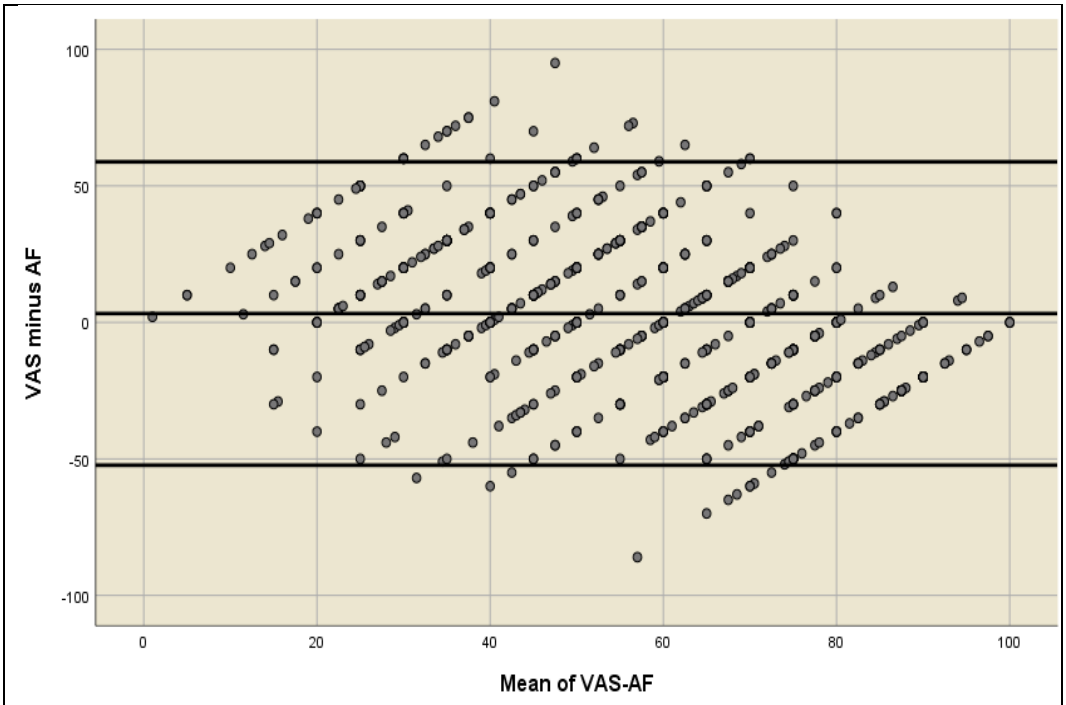


Figure 37 Bland Altman plot baseline: VAS and Angina Frequency
Mean = 3.24, Upper 95% CI is 59, Lower 95% CI is -52

Figure 38 is the twelve-month plot and shows a mean of -14.03 indicating a bias in the opposite direction from baseline, now in favour of AF. The 95% CI are slightly narrower.

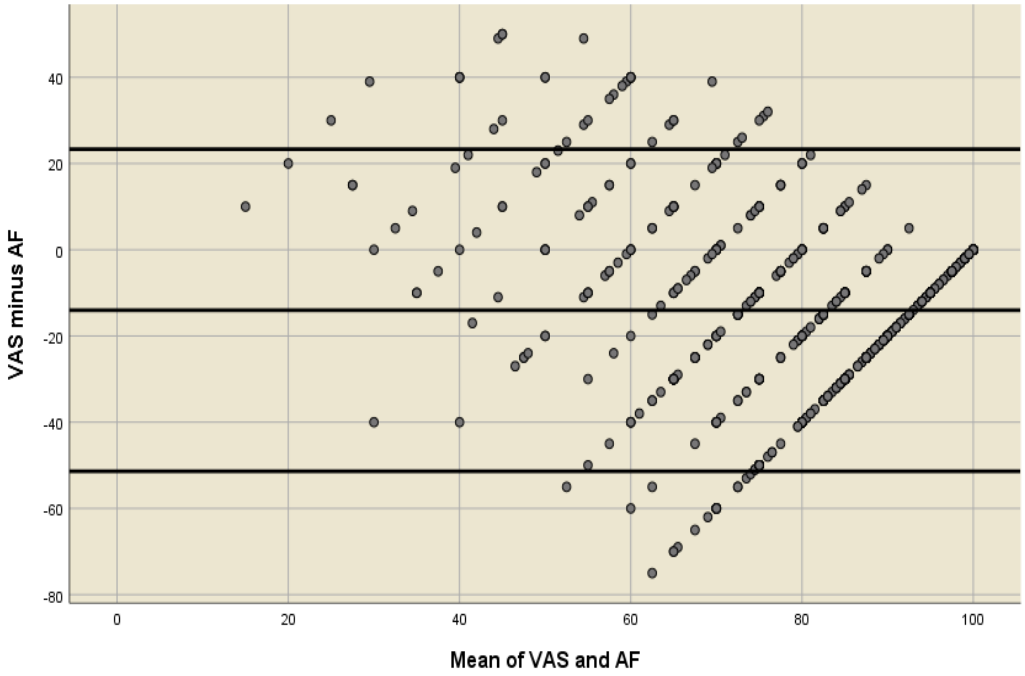


Figure 38 Bland Altman plot twelve-months: VAS and Angina Frequency
Mean = -14.03, Upper 95% CI = 23, Lower 95% CI = -51

5.14.2 Scatter Plots

Figure 39 is the baseline scatter plot and the correlation coefficient is 0.108 indication a poor relationship between VAS and AF.

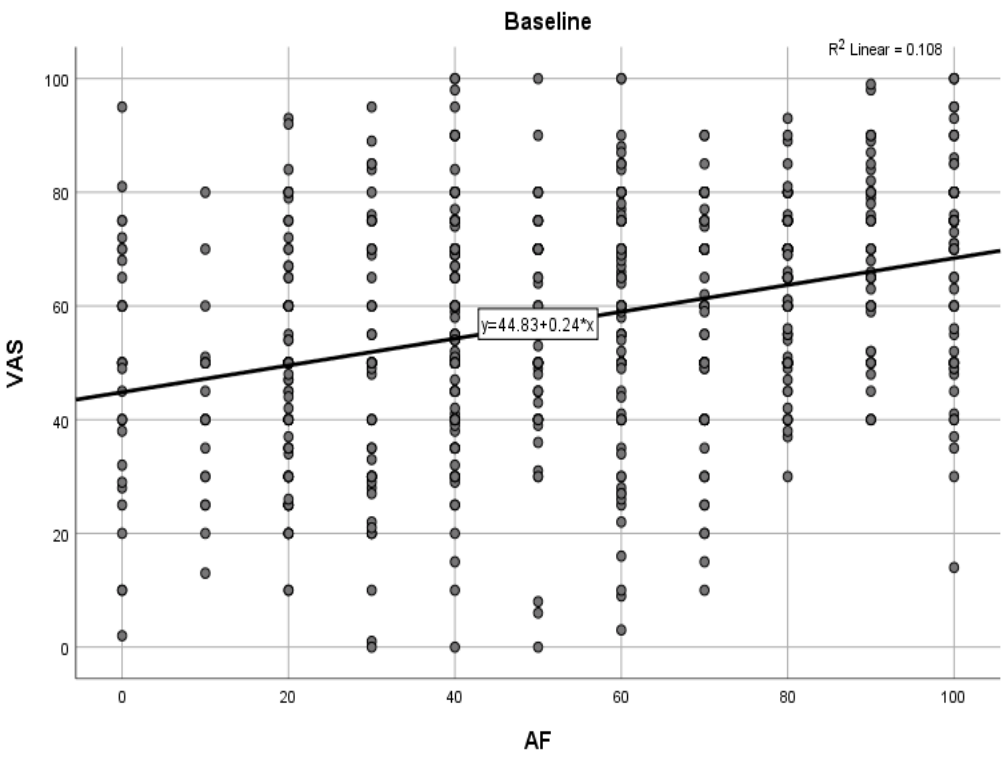


Figure 39 Scatter plot baseline: VAS and Angina Frequency
Correlation coefficient = 0.108

Figure 40 relates to twelve-months and shows a similar picture with a correlation coefficient of 0.181.

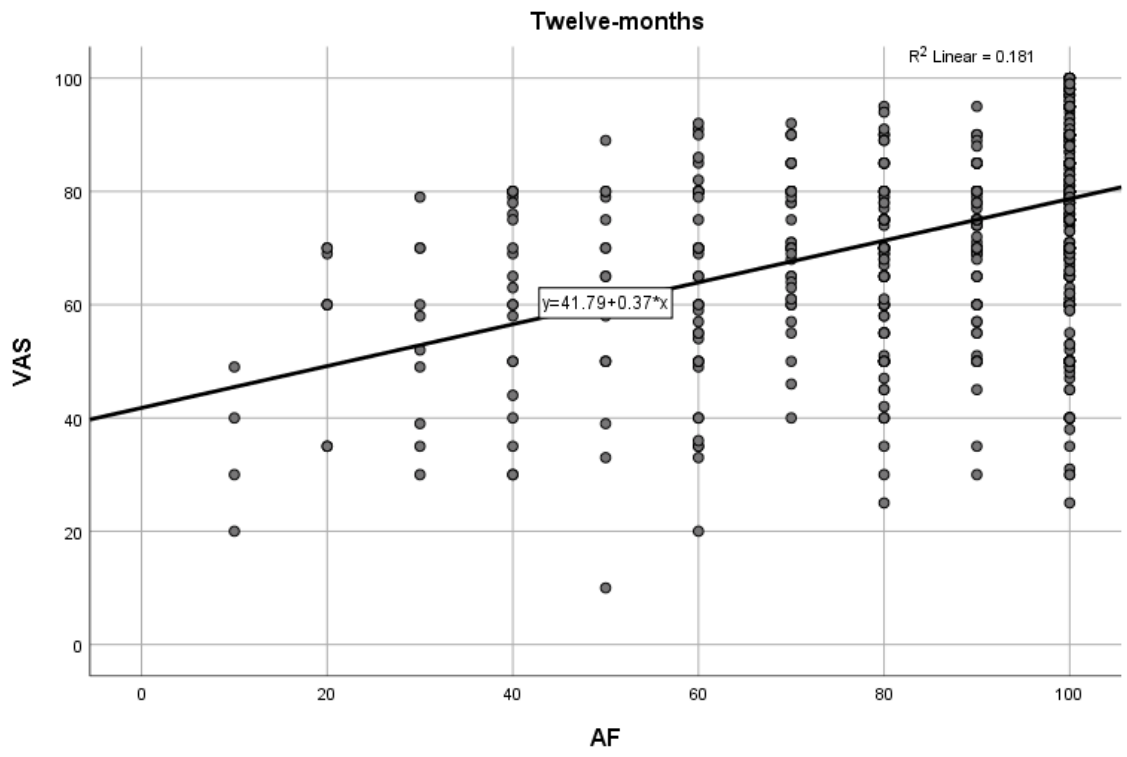


Figure 40 Scatter plot twelve-months: VAS and Angina Frequency
Correlation coefficient = 0.181

Figure 41 is the magnitude of change over twelve-months, and it also shows a poor correlation coefficient of 0.105.

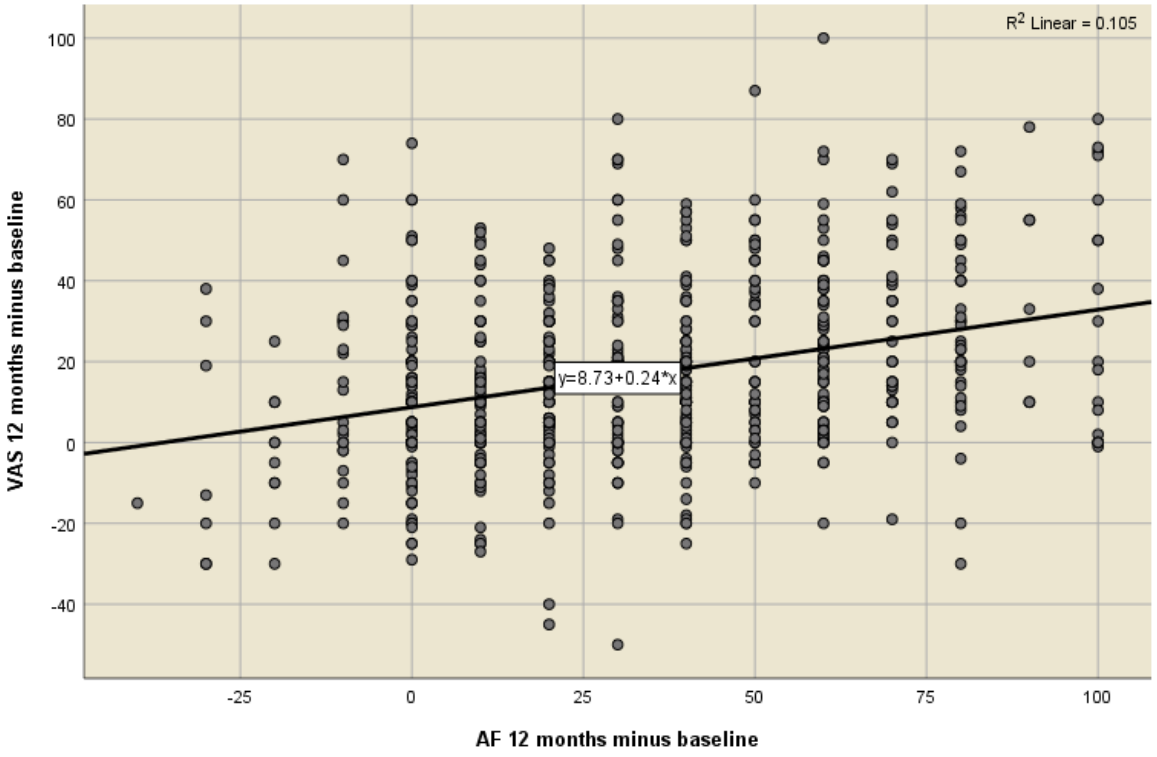


Figure 41 Change from baseline to 12 months: VAS and Angina Frequency
Correlation coefficient = 0.105

5.15 VAS and Treatment Satisfaction

5.15.1 Bland Altman

Figure 42 is the baseline plot. It shows a negative mean of -23.85 indicating a bias towards TS. The majority of data points on the 'x' axis are above 60 and the CI's are very wide.

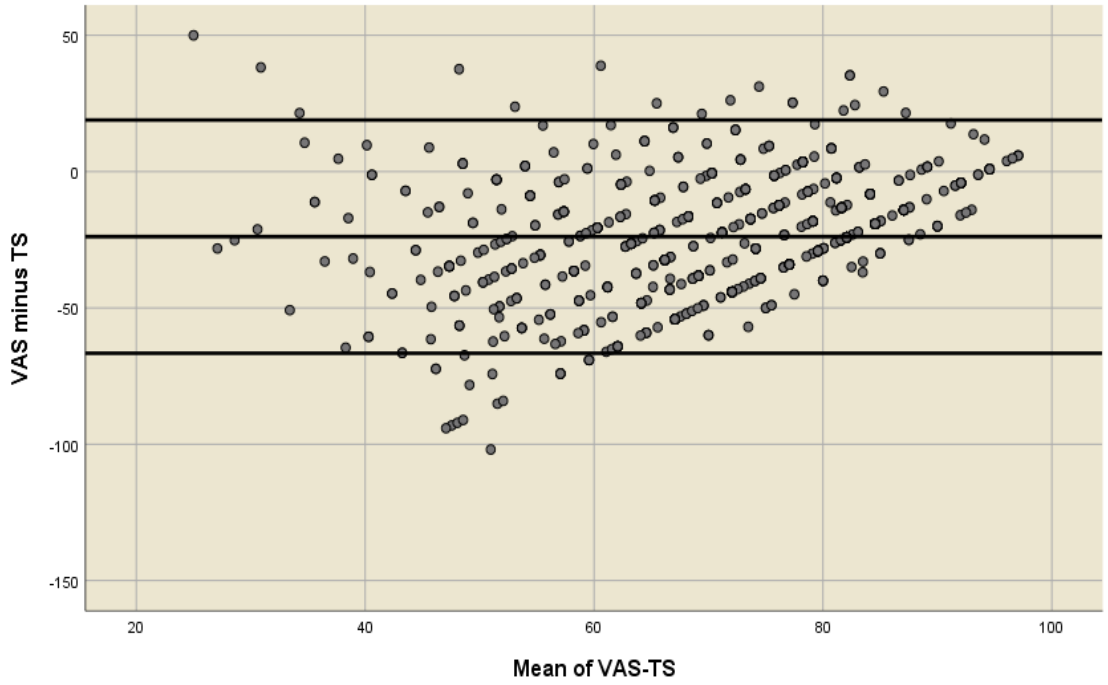


Figure 42 Bland Altman plot baseline: VAS and Treatment Satisfaction
Mean = -23.85, Upper 95% CI is 19, Lower 95% CI is -67

Figure 43 relates to twelve-months and again the mean is a negative -13.23 and a narrower 95% CI than baseline.

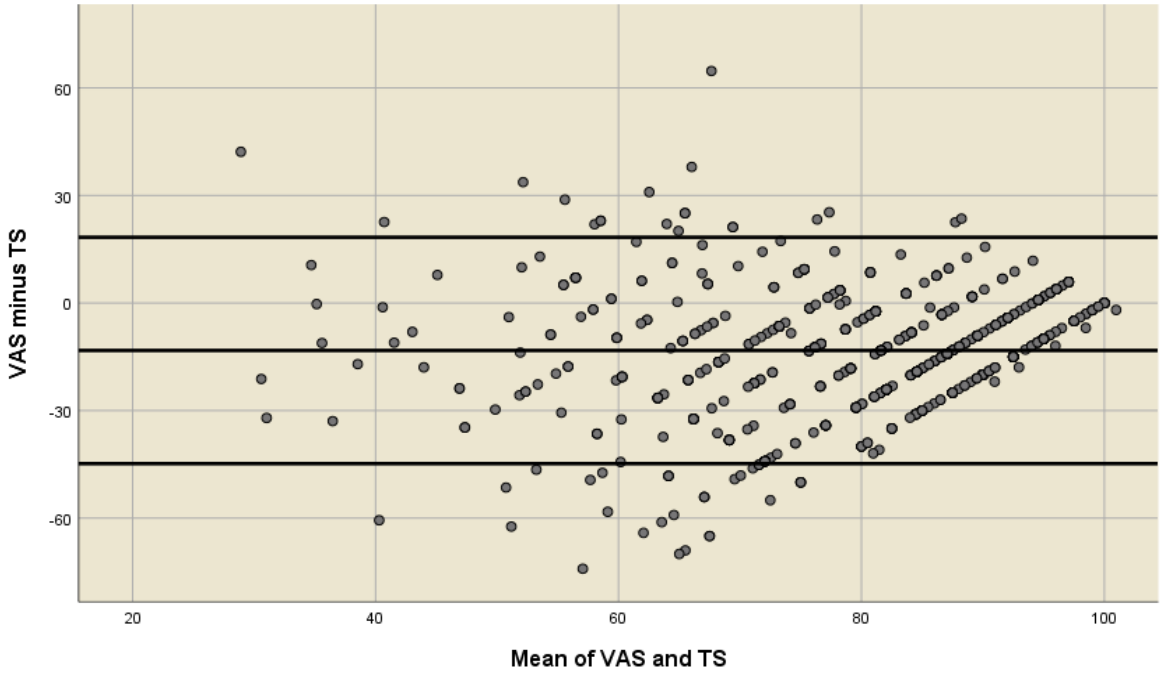


Figure 43 Bland Altman plot twelve-months: VAS and Treatment Satisfaction
Mean = -13.23, Upper 95% CI = 18, Lower 95% CI = -45

5.15.2 Scatter Plots

Figure 44 relates to baseline VAS and TS. The correlation coefficient is only 0.043.

There are a number of extreme outliers, such as VAS = 50, TS = 0 and VAS = 0 and TS = 100.

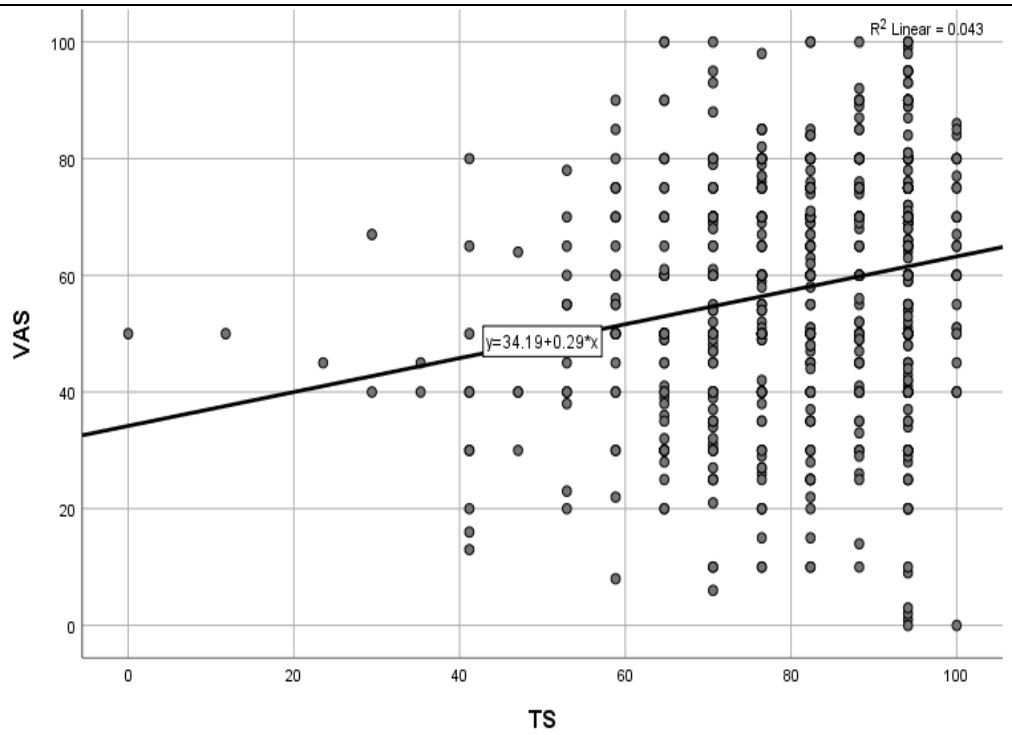


Figure 44 Scatter plot baseline: VAS and Treatment satisfaction
Correlation coefficient = 0.043

Figure 45 is at twelve-months and shows a slightly improved correlation coefficient of 0.203.

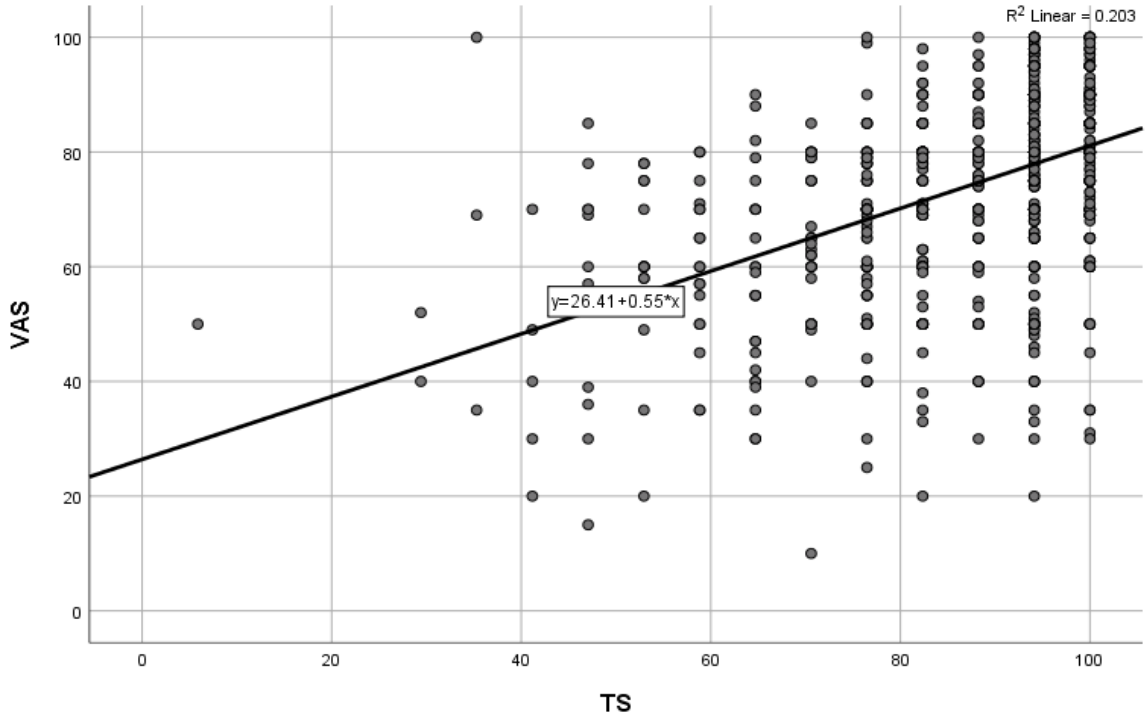


Figure 45 Scatter plot twelve-months: VAS and Treatment Satisfaction
Correlation coefficient = 0.203

Figure 46 is the magnitude of change from baseline to twelve-months comparing VAS with TS. The correlation coefficient is 0.03 indicating almost no relationship between the two axes.

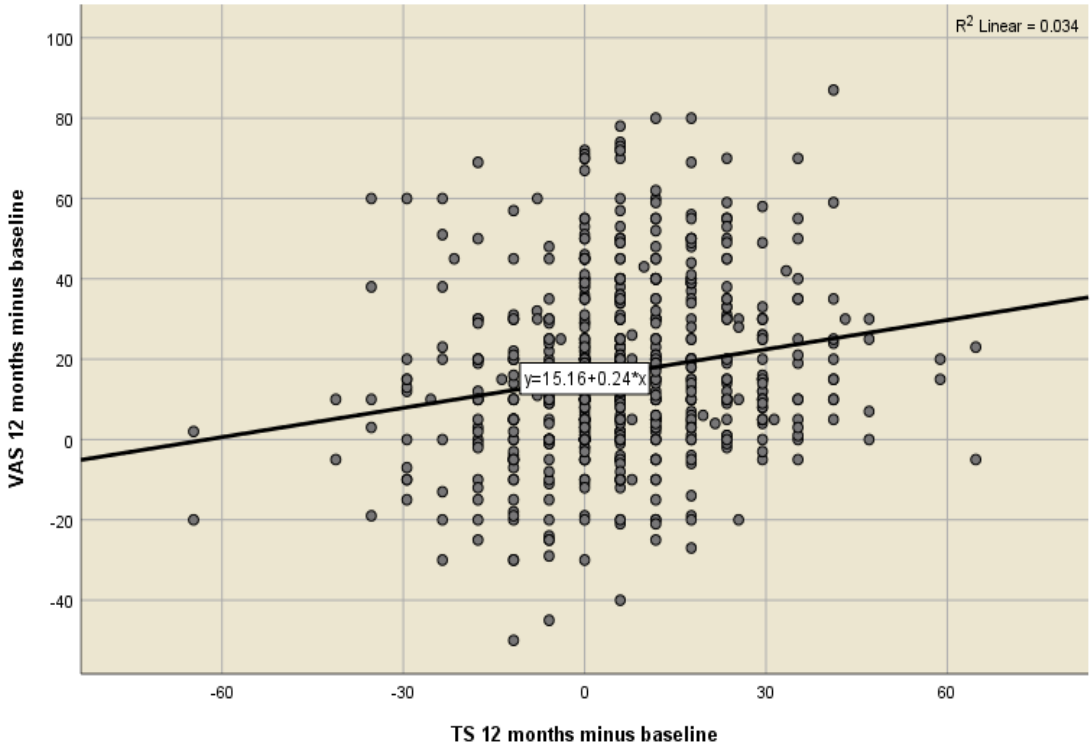


Figure 46 Change from baseline to 12 months: VAS and Treatment Satisfaction
Correlation coefficient = 0.034

5.16 VAS and Quality of Life

5.16.1 Bland Altman

Figure 47 is the baseline plot and the mean is 19.3 indicating a bias towards VAS.

There is a wide 95% CI of 60 to -21.

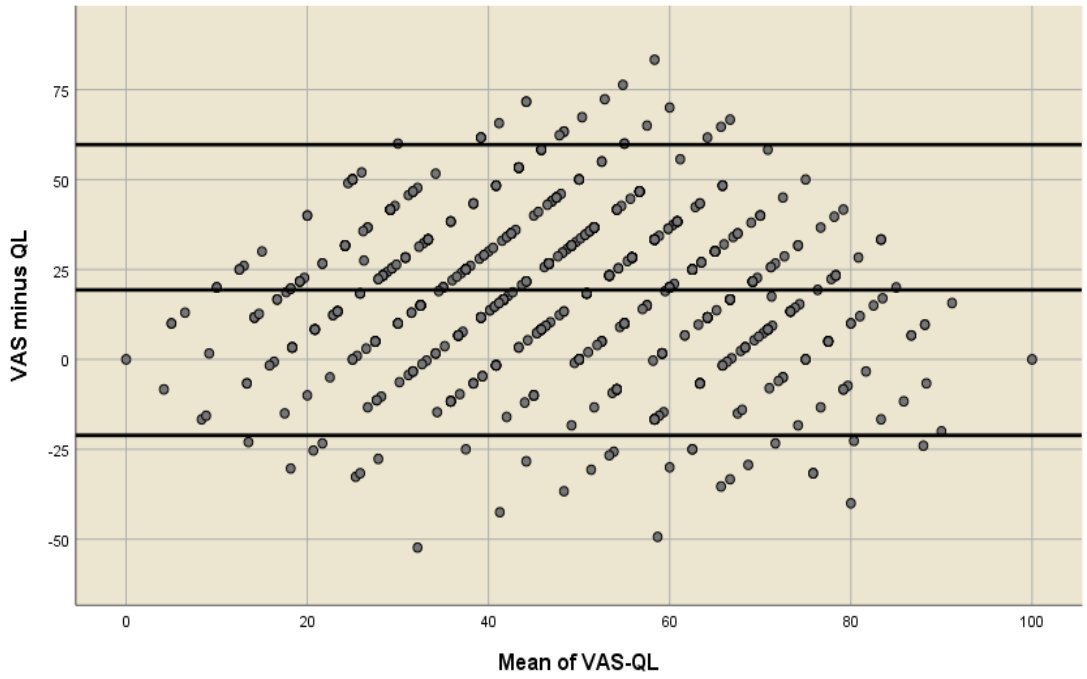


Figure 47 Bland Altman plot baseline: VAS and Quality of Life
Mean = 19.3, Upper 95% CI is 60, Lower 95% CI is -21

Figure 48 is for twelve-months, and the mean is now 2.28, so only a slight bias towards VAS. The 95% CI has shifted to 42 to -37.

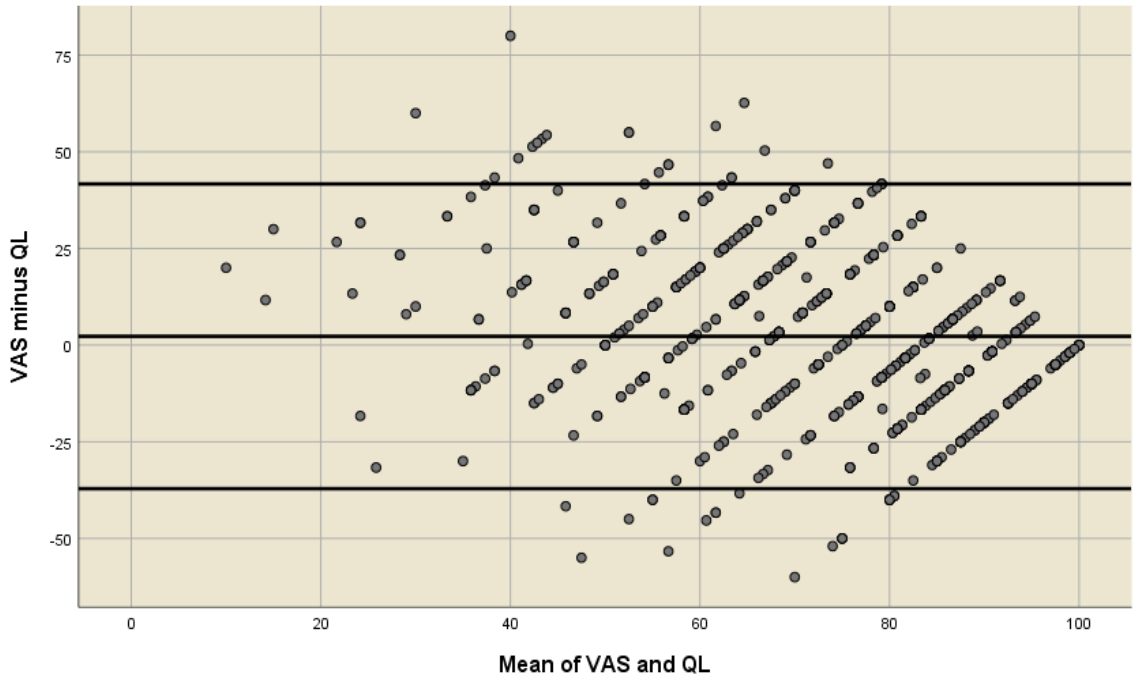


Figure 48 Bland Altman plot twelve-months: VAS and Quality of Life
Mean = 2.28, Upper 95% CI = 42, Lower 95% CI = -37

5.16.2 Scatter Plots

Figure 49 is baseline VAS and QL. There is a wide range of data points with a correlation coefficient of 0.229.

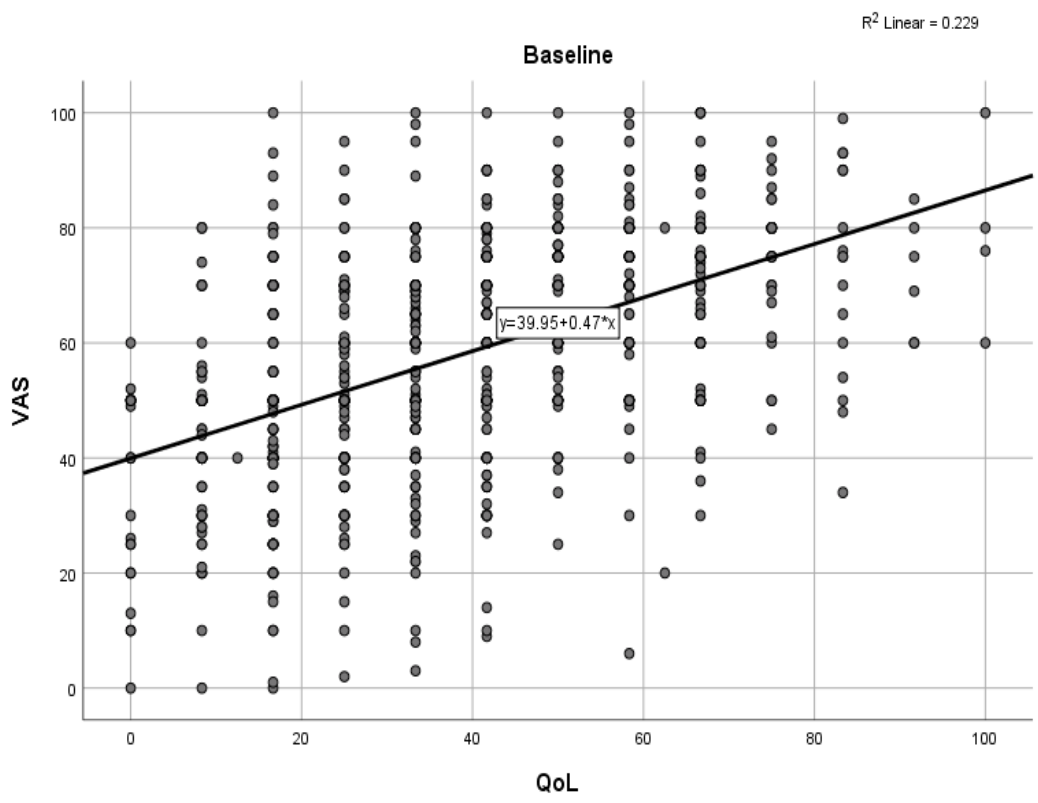


Figure 49 Scatter plot baseline: VAS and Quality of Life
Correlation coefficient = 0.229

Figure 50 shows an almost identical twelve-month correlation coefficient of 0.232.

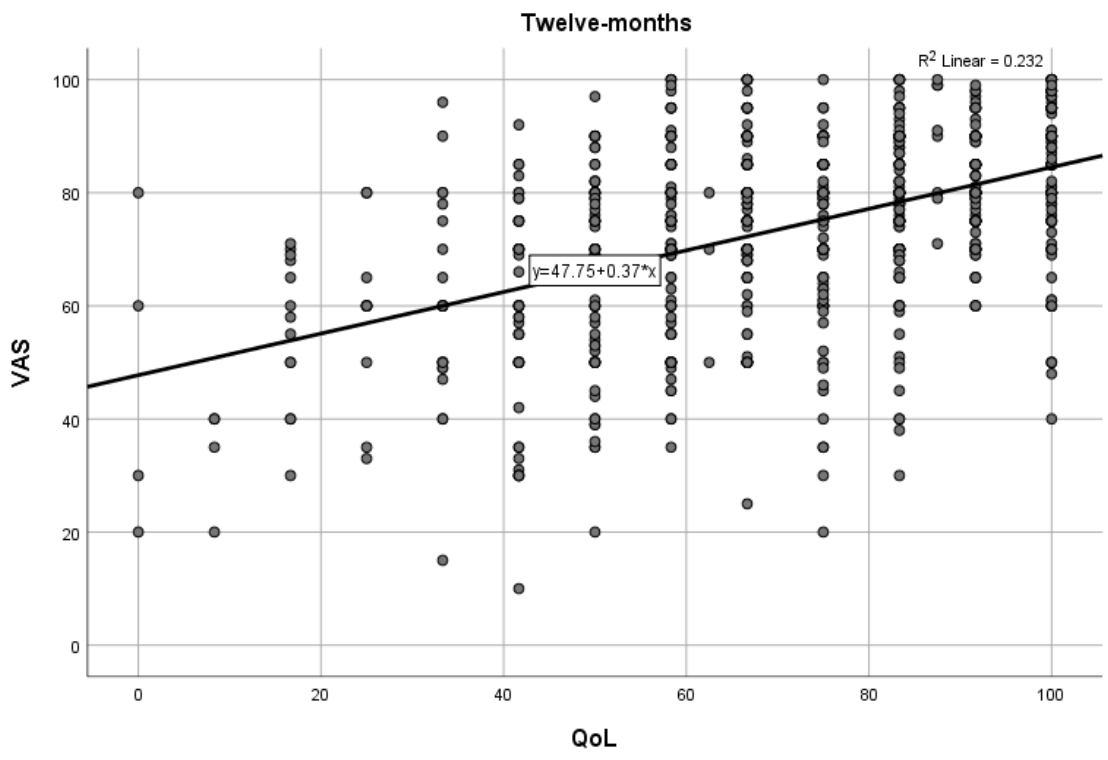


Figure 50 Scatter plot twelve-months: VAS and Quality of Life
Correlation coefficient = 0.232

Figure 51 is the magnitude of change from baseline to twelve-months comparing VAS with QL. The correlation coefficient is 0.144, a very poor relationship between the two axes.

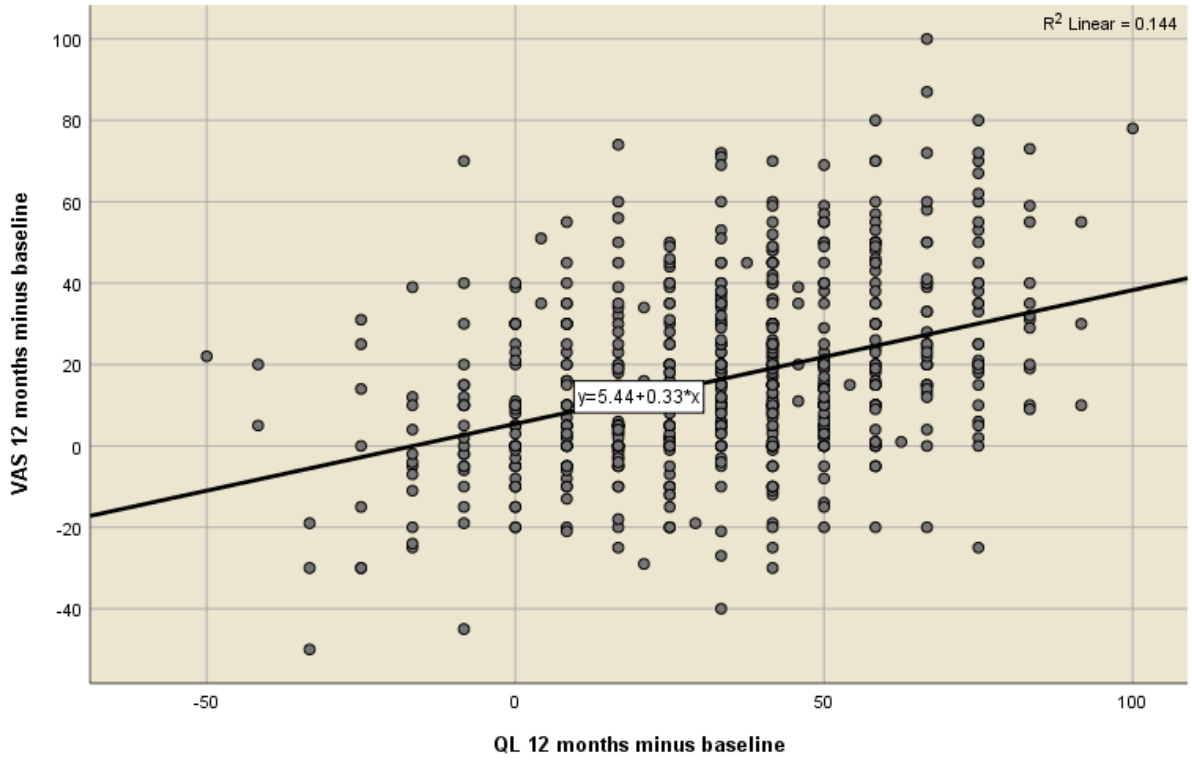


Figure 51 Change from baseline to 12 months: VAS and Quality of Life
Correlation coefficient = 0.144

5.17 Discussion

In the SoS trial, patients completed both the SAQ and EQ-5D. This created a unique opportunity to compare results from both instruments. As can be seen from the results above there is very poor relationship between the VAS and each domain of the SAQ.

The comparison in this chapter was undertaken to answer two fundamental questions. To what extent are the results at baseline and twelve-months similar? Secondly, how do the instruments compare in their ability to measure change from baseline to follow-up?

The Bland Altman plots relate to the VAS score minus each individual domain score at baseline. The plots were repeated using data from twelve-months. The confidence interval in the plots represents the 95% confidence around the mean. As can be seen for baseline figure 27 and figure 47, in every case the confidence intervals are wide. As for the sample size, the average numerator for the plots at baseline was 799 and at twelve-month it was 824. The take-away message from the Bland and Altman plots is there is a poor relationship between the two variables. The same is true of the scatter plots.

In a review of PROMs by Mackintosh et al in a report to the Department of Health in 2010, the recommendation for a disease specific instrument for angina was to use the SAQ. The rationale was that the SAQ had the most evidence supporting its use for patients undergoing CABG or PCI. The one drawback mentioned was the SAQ was developed for cardiac disease in general, not cardiac surgery [8]. It was stated that psychological well-being is a predictor of recovery and that a procedure such as CABG may have long-term effects on cognitive functioning. The SAQ does not include cognitive functioning questions.

5.18 Reasons for Discordance

There could be a number of explanations for the differences between the scores of VAS and SAQ domains. The first is that SAQ is a disease specific instrument and the other, EQ-5D is a generic instrument used in many different conditions and populations. They were developed with different intentions and, therefore, could not be expected to record identical elements of quality of life. The EQ-5D can be used in any population group. It can be used to calculate

cost-effectiveness or in a range of clinical trials in diverse disease populations. It can be used in healthy populations and between populations with different diseases [41]. It only contains five questions and a VAS. The SAQ by comparison was developed to quantify the physical and emotional aspects of coronary artery disease [30]. The SAQ contains nineteen questions in five domains.

There is still a discrepancy in what the patient is reporting between the SAQ and VAS. This is that the SAQ asks about how the patient felt over the previous *four weeks.* This means for the SAQ the patient has to consider their *average* in response to each question as opposed to how they feel when they complete the questionnaire. The VAS asks how the patient feels at that moment in time.

Another possible explanation for discordance is the effect of round number preference [102] affecting the VAS score but not the SAQ domain scores. This effect is seen in table 7 where, at baseline, 80% or the 819 respondents picked a VAS score that was divisible by 5. By contrast, table 8 shows that for the PL score only 11% of 815 respondents score was divisible by 5.

5.19 Administration of PROM

The method of administration of the PROMs may influence the results recorded by patients. Mercieca-Bebber et al [105] researched this, and found several issues in the way trial coordinators saw their role in the use of patient recorded outcomes. They refer to trial coordinators as either clinical research coordinator, site coordinator or research nurse. These staff are responsible for providing instructions to the participant in how to complete the questionnaire or PROM. This person also deals with enquiries from participants and may also enter data

into electronic systems such as a trial database. It is not known if each site participating in the SoS trial were given a Standard Operating Procedure (SOP) relating to the administration of the multiple PROMs. The authors state that although the type of PROM will be stated in trial protocols, details covering the method of administration are not often included. Although the same member of staff will give the same PROM to each patient, there may be subtle differences between multiple sites. In some sites, a member of staff may sit with the participant helping them to complete the PROM and answering questions. In another site a different approach may be taken where the member of staff leaves the room and lets the participant complete the PROM by themselves. The absence of a fully detailed standard method of PROM administration for every site may influence the results. However, it could be argued that as this was a randomised trial, any effects would be the same in each arm.

Research staff such as nurses and clinicians are very busy and can be involved with multiple studies simultaneously. This can have the effect of ensuring the PROM is completed in a standard way at the correct time-point a low priority. There are competing pressures which may contribute to the allocation of a limited time to the PROM. These can include managing trial governance, consenting participants, reporting adverse events, encouraging Principal Investigators to complete required paperwork, organising meetings and patient appointments, collecting specimens, managing budgets, liaising with sponsors and preparing ethics applications.

In amongst these varied duties, administration of a PROM may not be a high priority. Mercieca-Bebber et al [105] found that 85% of trial staff stated that patient reported outcome tasks formed a minor part of their role. Other issues identified included the organisational skills of trial staff, the level of empathy and rapport with participants, checking questionnaires for completeness and following up missing data and how to deal with data that was concerning in its nature. A significant challenge was balancing the needs of the patient with the needs of the trial. This was especially the case with participants who were very unwell or deteriorating when trial staff felt a flexible approach to data collection was warranted. Trial staff reported challenges with interpretation of the wording on questionnaires. The authors reported trial staff saying questions frequently missed by participants reflected a poorly worded question. Participants could challenge the need to answer a question that they felt was unrelated to their condition. The authors concluded that trial staff are at the heart of patient reported outcome data and their role is of the upmost importance. Further education was needed to ensure staff understand the impact of PROM administration.

With the SoS trial whatever shortcomings there may have been concerning how the PROM instruments were administered, the same issues would apply for patients allocated to each arm of the trial and between participating sites. This suggests there is something intrinsic about the PROMs themselves that accounts for the lack of correlation in scores.

5.20 Mode of Administration of PROMs

Bowling in 2005 [106] researched how the mode of questionnaire administration impacted the results. Modes of questionnaire delivery include person-to-person, participant completion alone in a clinical setting, postal delivery of questionnaires, electronic collection via website, telephone keypad or iPad etc. There are many influences on how participants may respond depending on the mode of administration. Thus, it can be difficult to separate out the effects of the mode of delivery on the results. Bowling states there are four steps involved in answering questionnaires which involve cognitive demands. These are understanding the question, recalling the information from memory, evaluating the link between the retrieved information and the question and finally communicating the response. The author states the most burdensome mode is likely to be self-administration as this demands the respondents are literate, do not have visual impairments and have dexterity of wrist and fingers. Respondents need the ability to tick boxes, to read and recognise numbers and write accurately and be able to follow instructions. The author also commented that the concept of "*data quality*" is a vague and there is no agreed "gold standard". It could be response rates, "accuracy" of responses (not defined) or absence of bias.

Non-measurement error was mentioned by Bowling as a potential reason for misleading results. This concerns the participants who are not included in a trial and how this may influence the results. The population chosen needs to be representative and from an up-to-date list. There can be sample selection bias

but in the SoS trial selection was limited by the trial protocol and participants who were eligible completed the same PROM instruments.

5.21 Influence of the Setting in PROM Completion

Bowling also states cultural, social and language may not have the same meaning to participants in questionnaires. The actual interaction can vary between settings, sessions and the member of staff conducting the collection process. Tourangeau et al [107] suggested a more subtle factor may be the pace of the interaction.

5.22 Conclusions

As seen from the results of the comparison between VAS scores and individual SAQ domains at both baseline and twelve-months, there is a very weak relationship. It is not, therefore, possible to recommend one or the other. Each instrument, the VAS and SAQ, is designed for a different purpose. The EQ-5D is calibrated for utility / cost analysis and can be used for many conditions and diverse populations. It is generic. By contrast, the SAQ is disease specific and aimed at patients with angina. It is designed to reflect the nuances of the quality of life for this patient population. The choice of instrument for a trial, therefore, depends on what outcomes are measured. If a primary outcome is a cost analysis then the EQ-5D is the obvious choice. If the outcome includes quality of life, then the SAQ is the instrument. There is nothing to stop both being used, other than an increased burden on the participant who will be required to complete both. Any researcher using both PROMs will need to be aware of the lack of

relationship between the two, either at a single time point or in terms of the magnitude of change.

The next chapter will examine if the same results can be obtained by comparing the full nineteen question SAQ with a shortened, seven question version.

Chapter 6 Validation of full SAQ with 7 question version

In this chapter I present an external validation of a truncated instrument - SAQ-7 - with data from the SoS trial using data recorded before and after revascularisation by CABG or PCI. The rationale for a shorter version of the SAQ is to make PROMs research easier, more practical, cheaper, more likely to be performed but without losing accuracy.

6.1 Introduction

The short form Seattle Angina Questionnaire (SAQ-7) has been designed to make research process more efficient, streamlined, acceptable, and cheaper to patients and researchers.

The full SAQ is designed to examine limitation in patients' quality of life and is, therefore, a disease specific PROM [30]. It consists of nineteen questions in five domains. The domains group questions in physical limitation (PL), angina stability (AS), angina frequency (AF), treatment satisfaction (TS) and quality of life (QoL). Traditionally, patients are asked to complete the SAQ prior to an intervention and again at a fixed time post-procedure. The two could then be compared to quantify the change in the patients' perception of their QoL. The SAQ has been endorsed by both the American College of Cardiology [56] and reviewed by the Patient-Reported Outcome Measurement Group, Oxford for the Department of Health [8].

There is strong support for the use of PROMs in clinical care but their use by clinicians in assessing their patient's QoL is rare [108]. One study found that none of the practices checked used the Kansas City Cardiomyopathy Questionnaire

(KCCQ) or another QoL measure and instead clinicians assessments of heart failure patients was the norm [108]. One author pointed out that key barriers to the universal adoption of PROMs in clinical practice could be summarised as the vowels: PROMs must be Actionable, Efficient, Interpretable, Obligatory and User-friendly. With this in mind, Chan et al developed a short version of the SAQ they called SAQ-7 [57]. This consisted of just seven questions in three domains. These were physical limitation, angina frequency and quality of life. Although the individual domains were scored from zero to one hundred (with higher scores indicating better QoL), a single overall score could also be calculated. This had the dual benefits of reducing the patient burden and providing clinicians and patients with a single score. The individual scores from SAQ-7 were compared to the SAQ and showed high levels of concordance (0.88 – 1.00). There was good construct validity, meaning SAQ-7 measured what it claimed to measure, and SAQ-7 was highly reproducible.

6.2 Literature Review

Details of the SAQ7 were first published in 2014 making a literature search less likely to return a high number of articles. The database used in the search was the Healthcare Database Advanced Search (HDAS). An iterative process was used which eventually returned four articles. The initial search was limited to "*Seattle Angina Questionnaire"* and "*Seattle Angina Questionnaire Short Version*", title only. The search string failed to return the UK version of the SAQ. The next iteration was *("Seattle Angina Questionnaire" AND ("short version" OR "SAQ7" OR comparison))*, *again title only*. This returned three results but still not the UK

version. Therefore, the final iteration wording for the search string was "'*Seattle Angina Questionnaire" and ("short version" or SAQ7 or SAQ-7 or SAQ-UK or comparison or compare)*. The criteria were title only and limited to articles in English. This returned four articles.

6.2.1 Garratt et al

Garratt et al developed what they called a UK version of the SAQ [109]. The authors rationale for the development of this United Kingdom version of the SAQ was to "*anglicise*" it and assess the measurement properties in patients with stable angina in the North of England. The instrument they developed was shorter than the original SAQ, having just fourteen questions. The assessment of the SAQ-UK was for reliability, validity and responsiveness. The authors reported that physiological measurements are poorly related to survival and health status. They point out the SAQ is one of the most widely used PROMs for angina patients but that it was validated using patients in the United States. The authors tested validity by comparing the SAQ-UK to the Sort Form SF-12 and EQ-5D. Data was collected in twelve general practices. There were 959 patients who received a questionnaire and of these, 655 (68%) took part in the study.

To "anglicise" the SAQ the word "*showering*" was changed to "*bathing or showering*". The phrase walking "*more than one block*" was changed to "*walking more than one-hundred yards*" and the word "*bothersome*" was changed to "*troublesome*". As further evidence to justify producing a shorter version of a PROM the authors mention the Short Form 36. This is a thirty-six-question generic PROM. However, the original eight domains in two groups of mental and

physical health were reduced to twelve questions used in both the United Kingdom and United States. The authors used Principal Component Analysis (PCA) [110] to assess the dimensionality of the instrument. The SAQ scores were compared to SF-12 and EQ-5D. the authors concluded their study did demonstrate a strong support for their modification of the instrument in United Kingdom patients. They also demonstrated that the five items removed had high levels of missing data and were not contributing to one of the clinically meaningful dimensions demonstrated by PCA. Following a further literature search for "SAQ-UK" in Pubmed in August 2020, no further articles were identified. This suggests the modified instrument has not been used since its development.

6.2.2 Chan et al

Chan et al developed a short, seven question version of the full SAQ [57] called the SAQ7. The rationale behind developing the SAQ7 was stated to be the limited use of the full SAQ in routine clinical practice. This was said to be in part due to its length, at nineteen questions. Chan et al used data from five longitudinal studies of coronary artery disease patients. This allowed data from 10,408 patients to be included from three distinct patient groups. These were stable coronary artery disease, elective PCI and finally acute myocardial infarction. The SAQ is divided into five domains, and for the SAQ7 the authors used the three that directly measure patients current health status. These are physical limitation, angina frequency and quality of life. The goal was to match comparability between the SAQ7 and SAQ.

Chan et al selected three items from the original nine item physical limitation domain. Each of the three items represented a different level of intensity. The three items were walking indoors on level ground, limitation of gardening, vacuuming or carrying groceries and finally lifting or moving heavy objects. Both items from the SAQ angina frequency were retained. From the original SAQ quality of life scale two questions were retained. These were enjoyment of life and feelings about spending the rest of your life with symptoms as they are now. Construct validity was assessed using Kendall τ-b rank correlation coefficient. The results for construct validity showed that there was good agreement between the SAQ7 and SAQ scores. With reproducibility, patients assessed between five and six months after elective PCI the SAQ7 showed excellent values and mean changes were less than 1 point. One month after PCI the mean SAQ7 scores increased by at least 18 points. The authors suggested this showed excellent reproducibility. Clinical change was reflected in the responsiveness scores. The authors pointed out that despite the SAQ being developed over twenty years ago the SAQ is seldom used in routine clinical care. Development of a short, validated version was an attempt to address this issue. A second issue addressed by the SAQ7 was to create a single "summary score" that combines the three domains. This, the authors suggested, may allow clinicians to quickly screen patients for a significant change in their health status [57].

6.2.3 Arnold et al

Arnold et al compared the SAQ with a daily angina diary [111]. Data was collected from 917 patients with stable angina and type II diabetes. The trial randomised

patients between placebo or ranolazine 1000mg twice a day. Trial subjects were given an electronic diary with built in prompts for daily entry. Subjects recorded angina episodes and use of sublingual medication. They also completed the SAQ at randomisation and again at eight weeks. In the analysis, patients were required to have taken the trial drug for at least 14 days and completed diary entries for at least 4 weeks. The SAQ angina frequency scores (SAQ-AF) were used in the comparison. The SAQ-AF frequencies were compared with the diary entries. The results were that there was a significant correlation between SAQ-AF and the diary entries of sublingual medication use. The authors suggested this was further support for the use of the SAQ in clinical trial of treatments designed to assess patients' angina symptoms.

The authors point out some limitations of the use of patient diaries. In studies using paper diaries, these can be completed inappropriately, for example by multiple entries in a single day. There can also be low compliance by patients. Electronic diaries are more reliable and have high rates of compliance. A downside is they are costly. Patients can still miss entering data for a day in an electronic diary. Electronic patient diaries are not practical in routine patient care. The authors also pointed out a limitation of the SAQ in comparison to daily diaries. This was that the SAQ asks about an average over the last four weeks. The SAQ also assumes that patients can in fact recall their angina status over the previous four weeks. The authors point out that "recall bias" is important with a symptom like angina which is episodic in nature. Recall bias is a tendency for selective memory of events from the past [112]. However, the correlations between diary and SAQ data suggests patients recall is not biased.

6.2.4 Dougherty et al

Dougherty et al compared three quality of life instruments [113]. These were the SAQ, SF-36 and Quality of Life Index-Cardiac VIII. The authors' rationale for the study was stated as coronary artery disease is often tackled for years before meaningful conclusions about quality of life are obtained. Medical care evolves over the years confounding the interpretation of outcomes. Quality of life measures provide a more responsive insight into health status. However, there are now multiple quality of life instruments available to be completed by patients. Three of these instruments were compared with the CCS classification to test reproducibility. Details about the Quality of Life Index can be found on this website https://qli.org.uic.edu/index.htm. The cohort was 107 patients.

The authors found with test-retest calculations all three instruments were stable with repeated administration at baseline and two-weeks. The SAQ scales had a close relationship with functional status defined by the CCS. The SF-36 did not demonstrate such a close relationship with the CCS and the Quality of Life Index had the lowest correlation. The authors concluded that the SAQ was the most responsive to both angina status and clinical change for trials of angina. The authors did refer to CCS as the "gold standard". However, it should be noted the CCS was designed to be completed by physicians after a discussion with their patients as an assessment of their patients' angina burden. CCS is not routinely completed by patients.

6.3 Methods

6.3.1 Comparing SAQ with SAQ-7

I decided to compare the SAQ with derived SAQ-7 using data from the Stent or Surgery Trial (SoS) [33]. Trial participants were also required to complete several PROMs including the SAQ at baseline and again at six and twelve-months post randomisation. The method of scoring the SAQ and deriving a single summary score is described in detail below and was developed by Chan et al [57]. Each domain was transposed to a score from zero to one hundred. A summary score - detailed below - (SAQ-SS) was then derived by taking the average of three domains, PL, AF and QoL. My comparison uses both the SAQ-SS and a summary score derived from the short SAQ, the SAQ7-SS, also detailed below.

6.3.2 Method of establishing a summary score (SAQ-SS)

The full SAQ derives a score for each of the five domains. Chan et al derived a method of calculating a single overall score which incorporates elements of three domains: physical limitation, angina frequency and quality of life. These three domains were selected as Chan et al stated they directly measure patients' current health status [57]. Each domain is dealt with slightly differently. Neither the TS or AS domains were used in the SAQ-SS.

6.3.2.1 Physical Limitation Score

This is the most complex of the domain scores to calculate. An individual item score of '6' is treated as missing. If five or more responses are missing, no score is computed; otherwise, missing responses are imputed as follows. Questions are grouped into three levels of exertional requirements. The lowest group includes

dressing, walking and showering (1a, 1b and 1c); the middle group includes climbing, gardening and walking more than a block (1d, 1e and 1f); the highest group includes running, lifting and sports (1g, 1h and 1i). Within each group, if one or two responses are missing, then assign each of them the average of the non-missing responses in that group. If all responses in the lowest or highest group are missing, then assign each of them the average of the middle-group responses. If all responses in the middle group are missing, then assign each of them the average of the means of the lowest and highest groups. The score is then calculated by taking the average of the nine responses and rescaling to 0-100, as follows:

$$SAQ - PL = 100 * [(average\ of\ questions\ 1a - 1i) - 1]/4$$

6.3.2.2 Angina Stability Score

This domain contains one question which can be scored from 1 to 6. A score of '6' is treated as '3' for the purposes of scoring. If the response is missing no score is computed. The score is calculated by rescaling as follows:

$$SAQ - AS = 100 * [(Q2) - 1]/4$$

6.3.2.3 Angina Frequency Score

The Angina Frequency score corresponds to questions 3 and 4. These can be scored between 1 and 6. In this domain an item score of '6' is included. If both questions are missing no score is computed. Otherwise the score is calculated by rescaling as follows:

$$SAQ - AF = 100 * [(average\ of\ Q3\ and\ Q4) - 1]/5$$

6.3.2.4 Treatment Satisfaction Score

The treatment satisfaction score corresponds to questions 5 to 8. Question 5 is treated differently to questions 6 to 8. Question 5 score of '6' is treated as a '5'. If three or more responses are missing , no score is computed. The score is calculated by rescaling as follows:

$$SAQ - TS = 100 * [(average\ of\ Q5\ to\ Q8) - 1]/4$$

6.3.2.5 Quality of Life Score

The quality of life score corresponds to questions 9 to 11. If two or more responses are missing, no score is computed. Otherwise, the score is calculated by taking the average of the non-missing responses and rescaling as follows:

$$SAQ - QL = 100 * [(average\ Q9\ to\ Q11) - 1]/4$$

6.3.3 SAQ Summary Score

The summary score integrates three domains into a single score. The domains are physical limitation, angina frequency and quality of life. If all three domain scores are missing, no summary score is computed. Otherwise, the summary score is calculated as the average of the non-missing domain scores as follows:

$$SAQ - SS = average\ of\ SAQ - PL, SAQ - AF\ and\ SAQ - QL$$

6.4 Rationale for single summary score

Chan et al used data from five longitudinal cohort studies to calculate a single summary score using three domains and seven questions [57]. All 10,408 patients had coronary artery disease (CAD) and were in five multi-centre registries. The three types of CAD in the registries were stable CAD, elective PCI and acute

myocardial infarction (MI). With each clinical setting, Chan et al conducted a series of analyses to evaluate construct validity, reproducibility, responsiveness and predictive validity of the short seven question SAQ7 and did the same for the full nineteen question SAQ. The three items selected from the original nine-question physical limitation domain represented low, moderate, and high-intensity activity.

6.5 Method of Calculating Short Form SAQ-7

The PL uses three questions from the SAQ. These are 1b (walking indoors on level ground), 1e (gardening vacuuming or carrying heavy groceries) and 1h (lifting or moving heavy objects). The AF uses two questions, 3 (over the last 4 weeks, how many times have you had chest pain, chest tightness or angina) and 4 (over the last 4 weeks, on average, how many times have you had to take nitroqlycerin for your chest pain, chest tightness or angina). The QoL uses two of the three questions, question 9 (over the last 4 weeks, how much has your chest pain, chest tightness or angina interfered with your enjoyment of life) and 10 (if you had to spend the rest of your life with your chest pain, chest tightness or angina the way it is right now, how would you feel about this). Each of the three domains was transposed into a score ranging from zero to one hundred.

6.5.1 SAQ7-Physical Limitation

This is the most complex of the three domains used in the SAQ7-SS. For purposes of scoring, a response of 6 is coded as missing. If two or more responses are missing, no score is computed; otherwise, missing responses are imputed as follows. If the response to Question 1b or Question 1h is missing, it is assigned

the response from Question 1e. If the response to Question 1e is missing, it is assigned the average of the responses to Questions 1b and 1h. The score is then calculated by taking the average of the three responses and rescaling to 0-100, as follows:

$$SAQ7 - PL = 100 * [(Average\ of\ Q1\text{b, 1e and 1h})\text{-}1]/4$$

6.5.2 SAQ7 Angina Frequency

AF score relates to questions 2 and 3 in the SAQ. If both questions have no response, no score can be calculated. Otherwise the score is calculated by taking the average of the non-missing responses and rescaling to 100 as follows:

$$SAQ7 - AF = 100 * [(Average\ of\ Q2\ and\ Q3) - 1]/5$$

6.5.3 SAQ7 Quality of Life

This was originally referred to as 'disease perception' and uses two questions, Q9 and Q10. If both questions have no response, no score can be calculated. Otherwise the score is calculated by taking the average of the non-missing responses and rescaling to 100 as follows:

$$SAQ7 - QL = 100 * [(Average\ of\ Q9\ and\ Q10) - 1]/4$$

If all three summary scores are missing it is not possible to calculate the summary score for SAQ7. Otherwise, the SAQ7-SS is the average of the SAQ7-PL, SAQ7-AF and SAQ7-QL. The SAQ-SS could then be compared to the SAQ7-SS.

6.6 Results

Statistical analyses were performed using SPSS v26. Data from all three time-points – baseline, six and twelve-months was initially analysed separately and then combined. Scatter plots were created for each timepoint.

6.6.1 SAQ and SAQSS – Scatter Plot, Baseline

Figure 52 shows a strong, linear positive correlation with few outliers. The correlation coefficient was 0.957. The majority of the data points fall between a score of 20 and 80.

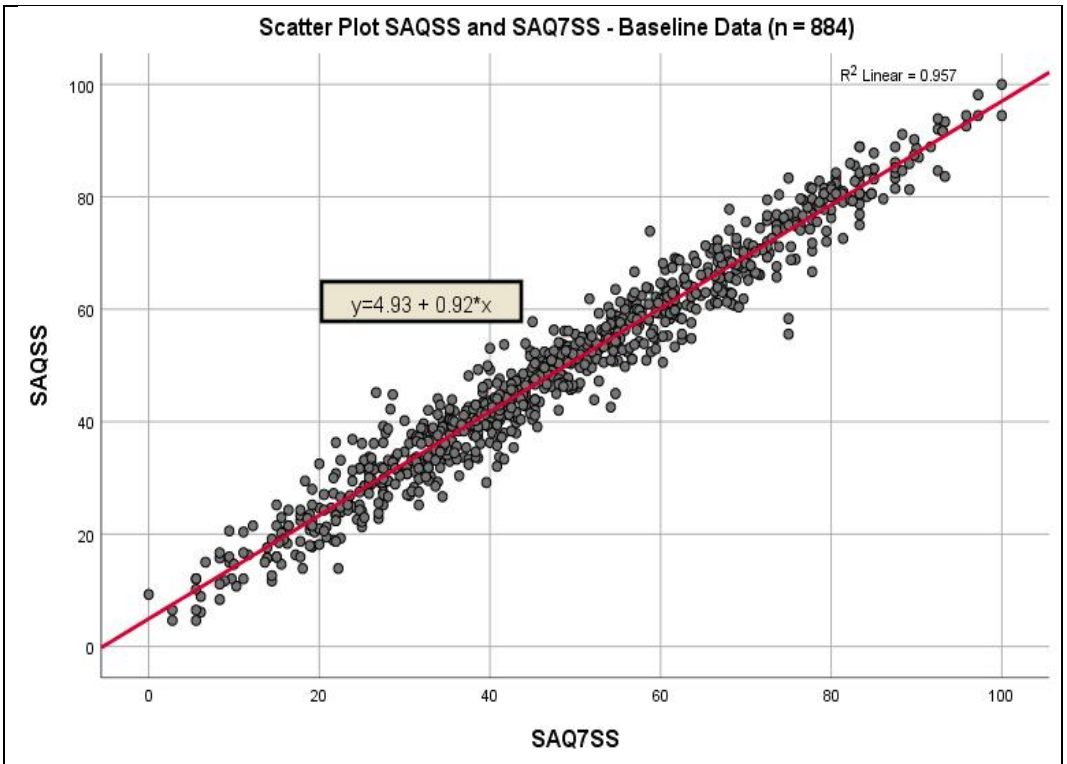


Figure 52 Scatter plot of SAQ and SAQ-7 summary score for baseline
Correlation coefficient = 0.957

6.6.2 SAQ and SAQSS – Scatter Plot, Six Months

Figure 53 is a scatter plot for six-months. Again, there is a strong, linear correlation. However, the majority of the data points are greater than 60 and the correlation coefficient is 0.951.

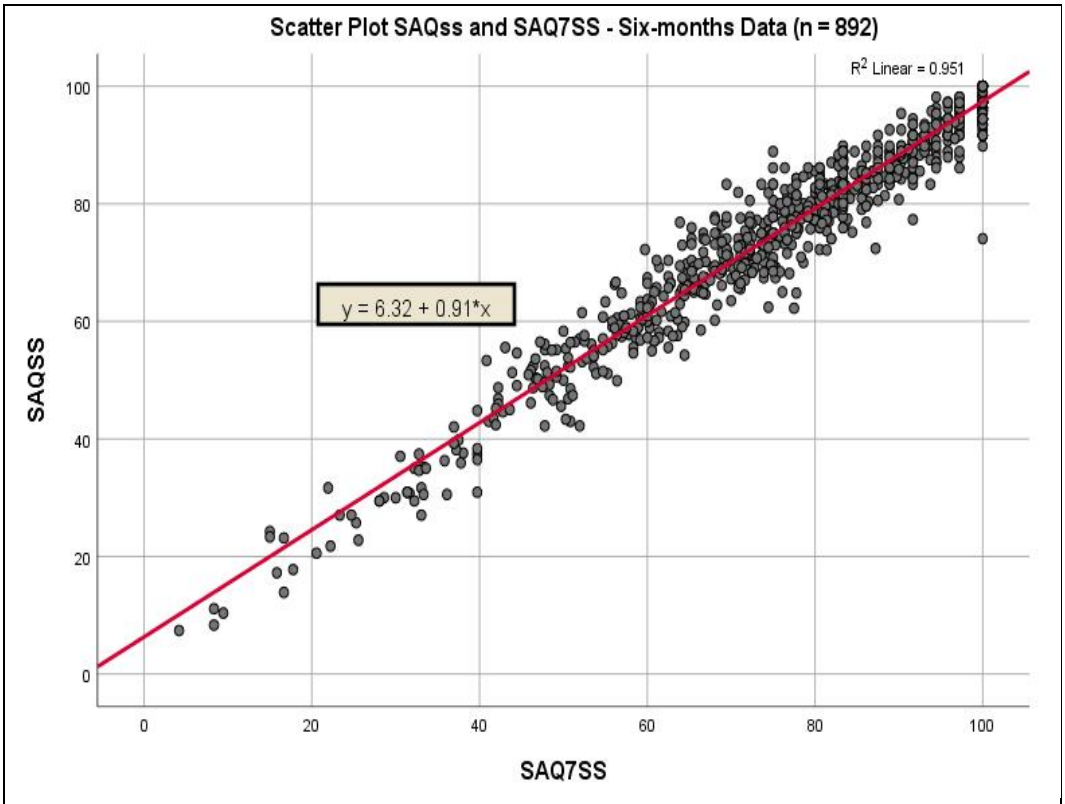


Figure 53 Scatter plot of SAQ and SAQ7 summary score for six-months

Correlation coefficient = 0.951

6.6.3 SAQ and SAQSS – Scatter Plot, Twelve-Months

Figure 54 is the scatter plot for twelve-months data and shows a correlation coefficient of 0.951. Two scatter plots, for six and twelve-months, display a homoscedastic pattern, and the six and twelve-month plots are characteristically fish shaped.

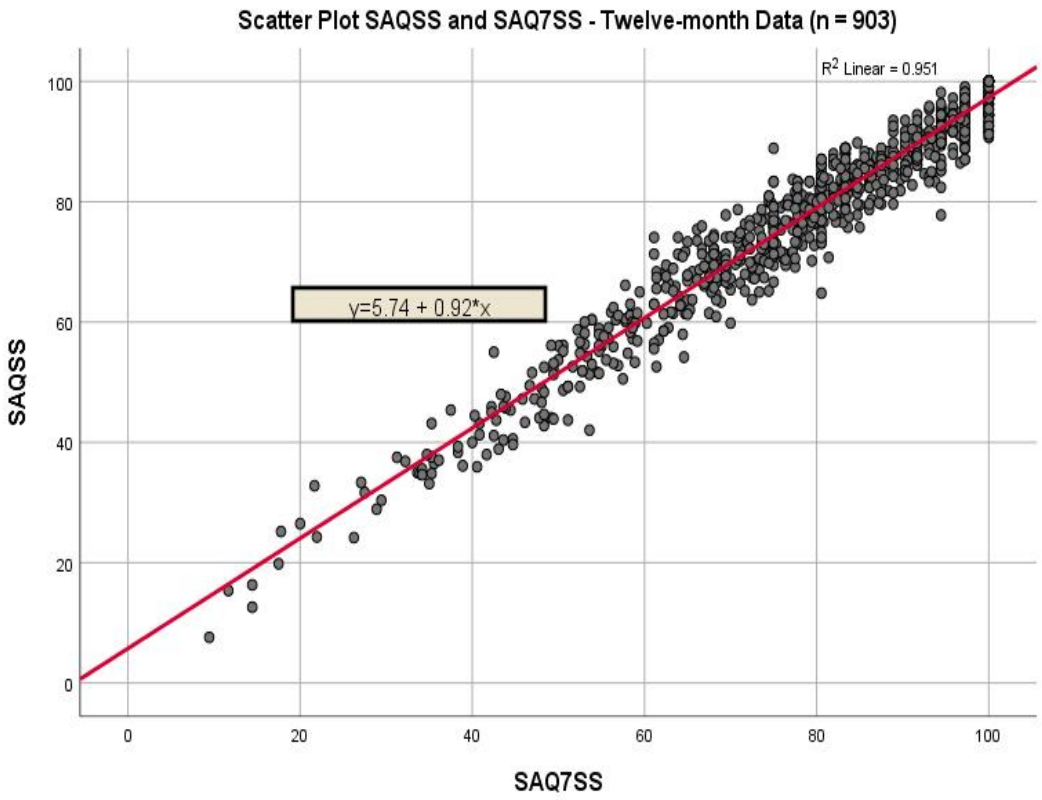


Figure 54 Scatter plot of SAQ and SAQ7 summary score for twelve-months

Correlation coefficient = 0.951

6.6.4 SAQ and SAQSS – Scatter Plot, All Three Time-points

Figure 55 combined all three time-points into a single scatter plot. The correlation coefficient is 0.97.

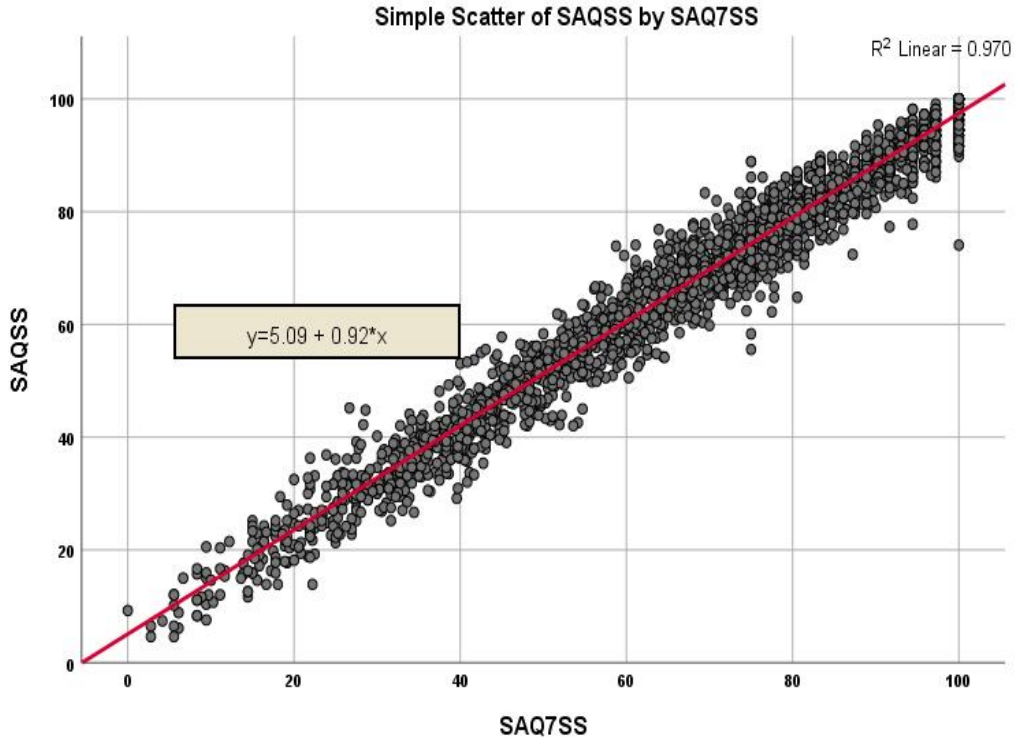


Figure 55 Scatter plot of SAQ and SAQ-7 summary score, all time-points

Correlation coefficient = 0.97

6.6.5 SAQ and SAQSS – Scatter Plot, CABG Patients

I also compared data relating to CABG and PCI but combining all time-points.

Figure 56 is data relating to CABG patients and shows a correlation coefficient of 0.967 and the majority of data points above 50. The numerator is 1365.

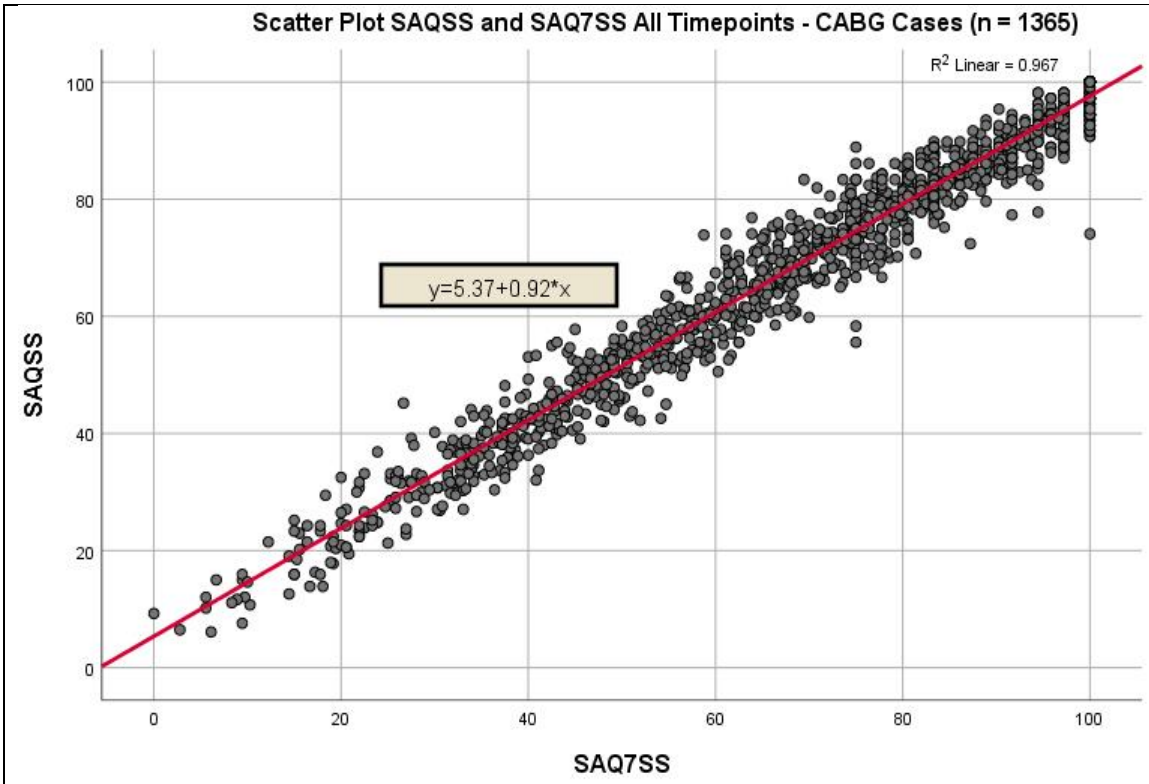


Figure 56 Scatter plot of SAQ and SAQ-7 summary score, CABG cases

Correlation coefficient = 0.967

6.6.6 SAQ and SAQSS – Scatter Plot, PCI Patients

Figure 57, shows a correlation coefficient of 0.973. With all scatter plots, there was always a very strong positive correlation between SAQSS and SAQ7SS. The correlation coefficient never went below 0.951.

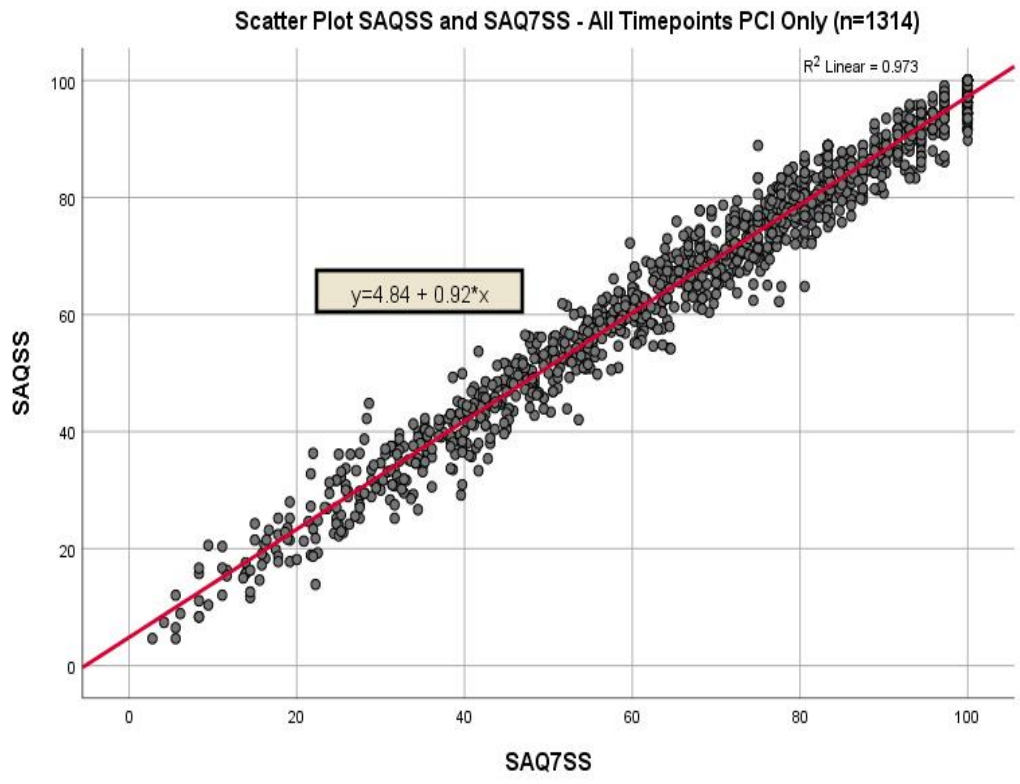


Figure 57 Scatter plot of SAQ and SAQ-7 summary score, PCI cases

Correlation coefficient = 0.973

6.6.7 SAQ and SAQSS – Bland Altman Mean Summary Score

I created a Bland Altman plot of the mean summary score by both SAQ7SS and SAQSS using all data from three time-points. (Figure 58). Positive differences indicate the SAQSS is greater than the SAQ-7, and negative differences the reverse. Ninety-five percent of most observations are between plus or minus 8 of the mean when the total score is one hundred. The mean itself is -0.16. In the mid-range between twenty to eighty there is good agreement. At low scores there is a tendency for SAQ-SS to report lower values. In the lower range, below twenty there are five cases with a mean greater than 8.36 (0.12% of the total), meaning the SAQ-SS was higher than the SAQ7-SS. However, in the higher range, above eighty, there are twenty-five cases with negative scores with a mean less

than minus 8.68 (0.93% of the total) indicating the SAQ7-SS scores more than the SAQ-SS.

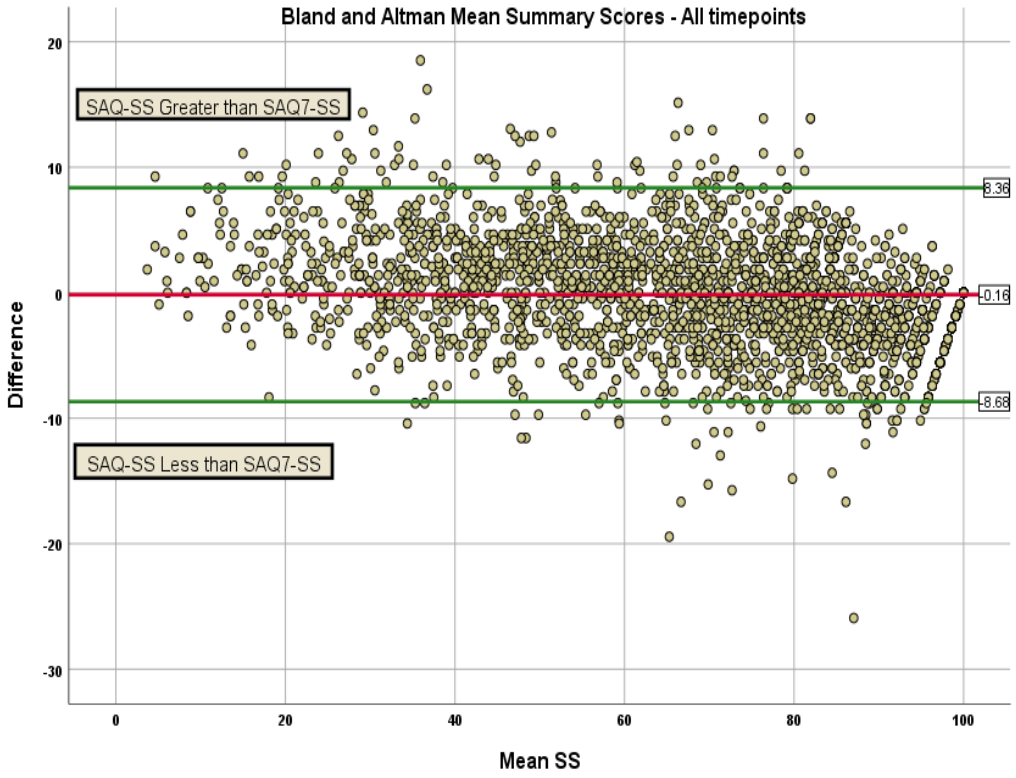


Figure 58 Bland and Altman Plot of mean summary score

95% CI = 8.36 to -8.68, mean = -0.16

6.7 Discussion

Outcome measures in studies examining the management of CAD have traditionally reported the incidence of adverse events such as mortality, or subsequent myocardial infarction or unplanned revascularisation. Important as these are, they tell us nothing about the patient's views of the effect of the intervention on their quality of their life. This is important as only a small percentage of patients die or experience adverse events in any specific period [(2)]. Even if used, physiological measurements do not directly correlate to health

status which is important to patients [109]. PROMs were developed to focus on the patients 'own assessment of their quality of life (QoL) and were routinely introduced into the National Health Service (NHS) in 2009 where they have an increasing role in clinical research[2]. PROMs are self-administered questionnaires that seek to quantify patients' symptoms, function and health-related quality of life [114].

There are limitations to the SAQ. One is the number of questions the patient is required to answer. Another is the absence of a single overall score which would make the PROM easier for both patients and clinicians to interpret. Garratt et al developed an 'anglicised' and shorter version, they called SAQ-UK [109]. The study used Principal Component Analysis and this resulted in five of the nineteen questions being removed from the questionnaire, leaving fourteen questions. These were grouped into three domains, Physical Limitation, Angina symptom and perception and finally treatment satisfaction. The SAQ-UK has been used in clinical trials [115] but no single, overall score can be calculated.

The 2007 report "Our NHS Our Future" [7] recommended routinely recording outcomes assessed by patients themselves. This was followed in 2008 by a requirement in acute services to report PROMs for four surgical procedures [26]. Since then PROMs have been collected for an ever-wider range of conditions and procedures. Some PROMs, such as the Short Form 36 contain thirty-six questions and are time-consuming for patients to complete. Others, such as the generic PROM EuroQol 5D are just five questions and are used widely. Two considerations for the more ubiquitous use of a PROM are the burden on

patients and the ease of interpretation by healthcare workers. The SAQ-7 meets both criteria with very high correlation between results from both SAQ-7 and SAQ.

The CCS is a tool used by clinicians to assess patient's angina burden. However, a recent study found discordance between clinicians and patients assessment of their angina burden when both used the CCS at the same timepoints [63]. This highlighted the need to pay more attention to the patient's assessment of their quality of life.

Black points out that the adoption of PROMs in England is largely driven by government wishes for the public to compare healthcare providers performance [20]. He also points out that there are three main challenges in implementing the widespread use of PROMs. These are minimising the time and cost of collection, analysis and presentation of data; achieving high rates of patient participation and thirdly, recognising the dimensions of quality, which are safety, effectiveness and experience [20].

6.8 Limitations

There are several limitations to this study. In creating a single summary score, we lose the discriminatory ability derived from specific domains. The SoS trial data is now over two decades old. It is conceivable that over that time-period patients' expectations and responses to the SAQ may have evolved. I have tested the SAQ-7 against a SAQ summary score that was itself 'created' by a methodology that may be flawed. The original SoS trial was limited to a cohort who were suitable for either PCI or CABG. This limited the population to a small percentage with

multi vessel disease. A further limitation was the reluctance of patients to consent to a random allocation of either PCI or CABG. The SoS trial authors speculated that only around 3-6% of patients undergoing multivessel revascularisation were in fact randomised [33]. The SoS trial was open label, and as such both patients and clinicians would be aware of the allocation. Therefore, it is conceivable this may have influenced decisions about symptom reporting. The SAQ does not ask about shortness of breath as the desire was to focus on coronary heart disease. Shortness of breath is a common comorbidity in patients with angina [116] and has a significant impact on quality of life.

6.9 Conclusions

The SAQ-7 can be completed by patients using a single side of paper in less than one minute. A summary score can easily be calculated providing healthcare workers with a single number between zero and one hundred. This incorporates elements of physical limitation, angina frequency and quality of life.

Chapter 7 examines the strengths and weaknesses of questionnaires.

Chapter 7 Questionnaires and Thesis Summary

7.1 Questionnaires: Quantitative Versus Qualitative Methods

Collating data collected in questionnaires uses a quantitative approach. Almalki [117] describes quantitative research as a "deductive approach" and that researchers regard the world as being outside of themselves. There is some "*objective reality independent of any observations*" [118]. Quantitative research uses mathematically based models and statistics to test a hypothesis. The use of questionnaires such as PROMs in quantitative research has several advantages alongside limitations. My research uses quantitative data derived from questionnaires. As such, every participant answers the same questions with identical multiple-choice response options. The questions themselves are what the researchers think of as the most important aspects of the subject. This may differ from what participants think of as important. This could be said to be detached from the idiosyncrasies which make people who they are, the minutiae of people's lives and their everyday reasons and explanations. It is beyond what LeCompte called "*vignettes and thick description*" [119]. There is no place for the nuanced detail of patients' perspectives in my thesis and I do not seek to claim this. The results from PROMs are simply aggregated into numbers. Quantitative and qualitative methods simply use a different lens to view the world. Quantitative may be considered a mile wide, qualitative is a mile deep. Quantitative data can be used to generalise about populations whereas qualitative research would not make this claim.

7.1.1 Advantages of Questionnaires

7.1.1.1 Financial Considerations

After paying for a licence to use a specific PROM, the PROM questionnaire itself can be used by a large number of respondents and, therefore, per respondent they are inexpensive. There is no requirement to hire specialist staff to deliver the PROM, it can be explained by existing research staff. The questionnaire could be available via a website or emailed to participants.

7.1.1.2 Practicality

A PROM can be targeted at a specific population and there is choice concerning the type of PROM which in turn dictates the type of questions involved. A PROM can be completed by large groups of a specific population in a short timeframe with minimum supervision or explanation.

7.1.1.3 Speed

A PROM can be a very quick way to obtain information about the participants. Researchers can avoid having to use any other organisation to deliver them and collect and collate results. It would be possible to collect information in a single day via a PROM completed online.

7.1.1.4 Ability to Scale Up

It is possible to scale up a questionnaire to include large numbers of respondents. It is possible to include a link to your website and this link could be in other specialised self-help websites. This link could invite responses from any or specified geographical regions. Questionnaires could be sent out by mass email or data collected from a convenience sample via a kiosk.

7.1.1.5 Comparability

A significant advantage of questionnaires is the ability to compare the summary of responses with other research using the same questionnaire. Alternatively, repeated questionnaires with the same population over time can be compared and statistically analysed.

7.1.1.6 Anonymity

A questionnaire completed and left in a local collection point may make the respondent feel anonymous, therefore potentially increasing completion rates. There is no way the investigator could contact the respondent to clarify a specific answer for example. However, a counter argument was put forward by Lelkes et al [120]. They argued that complete anonymity may in fact compromise accuracy. Complete anonymity reduces accountability. Their research found that completely anonymous questionnaires compromised accuracy rather than improving it. They referred to the process of "*satisficing*", a bias towards selecting the first response offered or selecting a "*Don't know*" option more often. It should be noted that in the SoS trial each participant was assigned a unique Case Record Number (CRN) and would be known to the staff at each participating site. The trial administrators would only see the CRN number and would not know the identity of the participant. It is not known what method was used in administering the PROMs to the participants, such as in person or by post.

7.1.1.7 Extensive Coverage of a Topic

A questionnaire can be extensive asking a full range of questions on a subject. These questions can cover several domains, including physical, emotional and

practical. It would also be possible in an online questionnaire to use branching logic. This means that the following question depends on the respondent's answer to the previous. This can lead to a great deal of details concerning one particular aspect of the respondent's life.

In summary of the advantages of questionnaires, they can be cheap, highly practical, fast, large scale, easy to compare to other questionnaires, anonymous and detailed.

7.1.2 Disadvantages of Questionnaires

Alongside the advantages of using questionnaires, there are disadvantages

7.1.2.1 Fatigue

O'Reilly-Shah [121] refers to respondent fatigue where respondents provide less thoughtful answers to questions in the later parts of a questionnaire. This can be caused by the length of the survey, the subject or the complexity of the questions such as the language used or if the question is open-ended or closed. A closed question may be "*Did you attend university?*" An open question might be "*How do you feel about attending university?*" This fatigue may lower the quality of the data or lead to an increase in unanswered questions.

7.1.2.2 Dishonesty

De Schrijver [122] pointed out that if a topic is of a sensitive nature, respondents may adopt one of three responses. The respondent may refuse to respond to the questionnaire at all, refuse to answer specific questions or answer questions dishonestly. Each will negatively influence data quality. Dishonest responses can be of two types. Desirable behaviour such as wearing a seat belt can be over

reported and undesirable behaviour such as drug use under reported. The author did state that classifying a question as "sensitive" was not straightforward. There is the potential for social pressure to affect a participant when dealing directly with a doctor. There may be a tendency for a participant to want to 'please' clinicians to avoid a 'fuss' by answering questions in a particular way.

7.1.2.3 Interpretation

Block [123] reported that in discussions, respondents often talked about issues raised in the questionnaire which yielded additional information which would not otherwise have been available to the researchers. The questionnaire represents a pre-conceived idea about the important aspects of a topic. However, when respondents talk about the topic, they raise issues not captured in the questionnaire. Interpretation is more difficult where a multiple-choice question is used. The respondent may have given a different answer had there been more choices available. Another problem is the interpretation of a scale giving a range of numbers. One person may respond with a six, and another a three, but this may mean the same to the individual respondents. However, analysis will treat the responses as different.

7.2 Bias in Questionnaire

The PROM most commented on in this thesis is the SAQ. This was evaluated for validity, responsiveness and reproducibility. However, the question of intrinsic bias invites discussion.

Choi and Pak state bias is a "*deviation of results or inferences from the truth, or processes leading to such a deviation*" [124]. They also state bias is a pervasive

problem in the design of questionnaires. They classify three main categories of bias. These are the way a question is written, the way the questionnaire is designed and the way the questionnaire is administered. Questions can be badly worded and ambiguous, too complex, too short, involve technical jargon or vague words. A scale on a questionnaire may be missing a “don’t know” option or there may be an overlapping interval. The authors also report bias as a result of the formatting of a question on a page, either vertically or horizontally. Central tendency bias occurs when a respondent selects the middle option from a list.

In an online blog, Jovancic [125] points out four types of bias in questionnaires. These are sampling, non-response, response and question order bias. Jovancic states sampling bias occurs as a result of the way respondents are selected. Response bias can be in the form of acquiescence bias, where the respondent tends to agree with whatever is being stated. Another form of response bias is demand bias. This occurs when the respondent attempts to guess what is behind the questionnaire and then respond with what they believe are the correct answers. Desirability bias occurs when the respondent provides answers that reinforce characteristics that are seen as socially desirable. Another form of bias is non-response. Even if the questionnaire is sent to all the relevant members of a population this does not guarantee they will all respond. Question order bias is where the initial questions could influence the following answers given. When there are multiple-choice answers, the order of these answer options can influence the results.

7.3 Key Thesis Findings

In chapter 4, I investigated discordance between clinician and patient on reporting QoL. I found no other references to both clinicians and patients reporting on the patients' CCS grade. There were key differences in the findings. Before the revascularisation procedure, clinicians tended to slightly differ from patients and reported a higher angina burden. However, at twelve-months this was reversed with a greater difference. It is possible patients may be reporting any chest pain as angina. Clinicians will be in a better position to attribute the pain to non-cardiac causes. The key implication is that studies reporting outcomes using CCS assessed by clinicians may overstate the effectiveness of treatment.

Chapter 5 examined the relationship between SAQ and EQ-5D VAS scores. I found the scatter plots showed a very poor correlation coefficient and the magnitude of change over twelve-months also differed and showed a wide standard deviation. EQ-5D and SAQ are different PROMs created for two distinct purposes. The EQ-5D VAS asks the patient how they feel now, whereas the SAQ relates to the previous four weeks. As there is such poor correlation between VAS and SAQ it would not be advisable to rely on just one of the PROMs.

Chapter 6 investigated if a shorter, seven question version of the SAQ would compare with the full SAQ. In each case a single summary score was created. Using a series of scatter plots I found the correlation coefficient was very high, typically above 0.95, almost a perfect correlation. However, it should be noted the summary score was created by a method that may contain flaws. Using just seven questions does lose some of the discriminatory ability of the full SAQ.

However, a short disease-specific PROM with a single summary score may improve the take-up of this instrument in day to day clinical practice and provide useful information for both clinicians and patients.

Chapter 7 examined issues abound the use of questionnaires. There are several advantages and disadvantages. Gathering large amounts of data from multiple patients will often involve a questionnaire. This occurs not just in the healthcare setting but in all walks of life. There will always be limitations and biases in the design, layout and population invited to participate. However, questionnaires remain a key element of PROMs. A greater understanding of their limitations and potential biases would be an advantage for any researcher.

7.4 Routine Clinical Use of PROM Data

Black [20] pointed out that routine use of PROMs is restricted to England, Sweden and some parts of USA. One further observation by Black was that in England the use of PROMs data has been driven by government wishes for the public to compare their healthcare providers performance. It is substantive. However, in Sweden and the USA the medical profession has focused on improving clinical care of individual patients, a formative approach. This represents a fundamental difference in the use of PROMs data.

If the fundamental aims of healthcare are to improve patients' QoL then patients themselves are in the best position to report on this. As Black points out, patients do welcome being involved in decisions about their care and treatment. A response rate for a patient who only completes one questionnaire is likely to be higher than for a clinician who has to complete one for every patient. Asking a

clinician to assess their own practice may introduce observer bias. Observer bias is defined as "*Systematic difference between a true value and the value actually observed due to observer variation*" [126]. Finally, ensuring patients' views are at least considered in strategic decision making is likely to influence public accountability.

Griggs et al [21] referred to PROMs as when maximised in clinical settings can "*be leveraged to inform clinical decision making, to improve quality of care and to foster communication between patients and providers*". The authors also went on to state PROMs "*have the potential to be as valuable to the clinical encounter as a stethoscope is to the physical examination"*. However, the large-scale use of PROMs data requires both local infrastructure and local hospital Trust support. This can be costly and complex. To calculate a score for SAQ data I developed a spreadsheet in Microsoft Excel which used multiple columns and formulas. This could be built into hospital Trust Electronic Patient Record (EPR) systems. However, the SAQ is just one of many PROMs used and is specific to ischaemic heart disease. My Trust also treats patients with heart failure, valve disease, lung cancer, cystic fibrosis, congenital heart disease and those with arrhythmias needing pacemakers. Each population have different needs and utilise different PROMs. Building automatic systems to record PROMs for each disease population is possible but would require a commitment from the executive team, a dedicated development team including education and Information Technology and sufficient long- term funding.

Patients need to understand the potential of PROMs and perhaps local patient groups may help in this endeavour.

7.5 An Opportunity for Improved Care

This thesis could be said to reflect the famous quote from Dr William Osler to "Listen to your patient" "He's telling you the diagnosis" [127]. Osler was a pioneer, one of the first physicians to take his students out of the lecture theatre onto wards to meet patients. The NHS needs new pioneers. Healthcare is perhaps a misnomer as the NHS is predominately concerned with disease management and this is a costly business. In 2018 the healthcare expenditure was £214.4 billion, or £3,227 per person [128]. That equates to 10% of this country's gross domestic product. Total healthcare spending has doubled in the period 1997 to 2018.

There have been great strides in incorporating routine collection of PROMs data in some surgical procedures such as hip and knee replacement. However, this data is summative and can be used to compare results between trusts. A formative use of PROMs could inform both the patient and physician and aid decision making that benefits the patients' QoL.

7.6 Key Conclusions

My work in chapter 6 comparing a short version of the SAQ containing just seven questions showed a very close relationship between the SAQ-7 and full SAQ summary score. It is possible to shorten a well-established and commonly used PROM. Although calculating the summary score involves several stages, this could be incorporated into an EPR system and used routinely. A summary score is easier to interpret than 5 individual domain scores, making sequential scores

easy to present in a graph for example. This information could then be seen by both clinicians and patients at each consultation or on ward rounds.

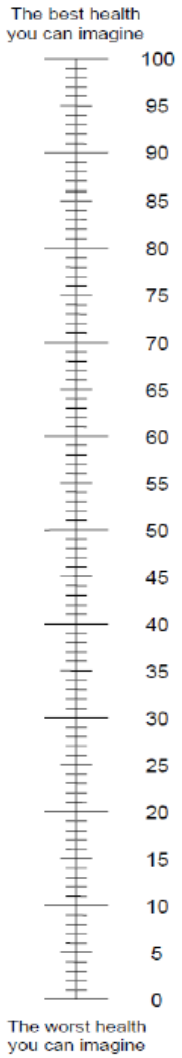
This would be a small but important way to begin the process of changing the culture of PROMs in the UK. It would create the environment to incorporate both the clinicians' expertise and patients' QoL derived from a PROM. Ultimately, physicians need to embrace the potential of PROMs and, as Griggs et al state "*achieve the dual goals of value based and patient centred care*" [21].

Appendix i Example of EQ-5D

Domain	Level	Tick if 'Yes'
Mobility	I have no problems in walking about	
	I have some problems in walking about	
	I have extreme problems in walking about	
Self-care	I have no problems washing or dressing myself	
	I have some problems washing or dressing myself	
	I have extreme problems in washing or dressing myself	
Usual activities	I have no problems with doing my usual activities	
	I have some problems with doing my usual activities	
	I have extreme problems doing my usual activities	
Pain Discomfort	I have no pain or discomfort	
	I have some pain or discomfort	
	I have extreme pain or discomfort	
Anxiety Depression	I am not anxious or depressed	
	I have some anxiety or depression	
	I have extreme anxiety or depression	

- We would like to know how good or bad your health is TODAY.
- This scale is numbered from 0 to 100.
- 100 means the best health you can imagine.
  0 means the worst health you can imagine.
- Mark an X on the scale to indicate how your health is TODAY.
- Now, please write the number you marked on the scale in the box below.

YOUR HEALTH TODAY = ☐



3

Appendix ii Example of SAQ

The Seattle Angina Questionnaire

1. The following is a list of activities that people often do during the week. Although for some people with several medical problems it is difficult to determine what it is that limits them, please go over the activities listed below and indicate how much limitation you have had **due to chest pain, chest tightness, or angina over the past 4 weeks.**

Place an x in one box on each line.

Activity	Severely Limited	Moderately Limited	Somewhat Limited	A Little Limited	Not Limited	Limited, or did not do for other reasons
Dressing yourself	☐	☐	☐	☐	☐	☐
Walking indoors on level ground	☐	☐	☐	☐	☐	☐
Showering	☐	☐	☐	☐	☐	☐
Climbing a hill or a flight of stairs without stopping	☐	☐	☐	☐	☐	☐
Gardening, vacuuming, or carrying groceries	☐	☐	☐	☐	☐	☐
Walking more than a block at a brisk pace	☐	☐	☐	☐	☐	☐
Running or jogging	☐	☐	☐	☐	☐	☐
Lifting or moving heavy objects (e.g. furniture, children)	☐	☐	☐	☐	☐	☐
Participating in strenuous sports (e.g. swimming, tennis)	☐	☐	☐	☐	☐	☐

2. Compared with 4 weeks ago, how often do you have **chest pain, chest tightness, or angina** when doing your **most strenuous** level of activity?
I have had **chest pain, chest tightness,** or **angina**...

Much more often	**Slightly more** often	About **the same**	**Slightly less** often	**Much less** often
☐	☐	☐	☐	☐

3. Over the past 4 weeks, on average, how many times have you had **chest pain, chest tightness, or angina?**
I get **chest pain, chest tightness,** or **angina**...

4 or more times per day	1-3 times per day	3 or more times per week but not every day	1-2 times per week	Less than once a week	None over the past 4 weeks
☐	☐	☐	☐	☐	☐

4. Over the past 4 weeks, on average, how many times have you had to take nitros (nitroglycerin tablets) for your **chest pain, chest tightness,** or **angina?**
I take nitros...

4 or more times per day	1-3 times per day	3 or more times per week but not every day	1-2 times per week	Less than once a week	None over the past 4 weeks
☐	☐	☐	☐	☐	☐

5. How bothersome is it for you to take your pills for **chest pain, chest tightness or angina** as prescribed?

Very bothersome	**Moderately** bothersome	**Somewhat** bothersome	**A little** bothersome	**Not** bothersome at all	My doctor has **not prescribed** pills
☐	☐	☐	☐	☐	☐

6. How satisfied are you that everything possible is being done to treat your chest **pain, chest tightness, or angina**?

Not satisfied at all	Mostly dissatisfied	Somewhat satisfied	Mostly satisfied	Highly satisfied
☐	☐	☐	☐	☐

7. How satisfied are you with the explanations your doctor has given you about your **chest pain, chest tightness, or angina**?

Not satisfied at all	Mostly dissatisfied	Somewhat satisfied	Mostly satisfied	Highly satisfied
☐	☐	☐	☐	☐

8. Overall, how satisfied are you with the current treatment of your **chest pain, chest tightness, or angina**?

Not satisfied at all	Mostly dissatisfied	Somewhat satisfied	Mostly satisfied	Highly satisfied
☐	☐	☐	☐	☐

9. Over the past 4 weeks, how much has your **chest pain, chest tightness, or angina** interfered with your enjoyment of life?

It has **severely** limited my enjoyment of life	It has **moderately** limited my enjoyment of life	It has **slightly** limited my enjoyment of life	It has **barely** limited my enjoyment of life	It has **not** limited my enjoyment of life
☐	☐	☐	☐	☐

10. If you had to spend the rest of your life with your **chest pain, chest tightness, or angina** the way it is right now, how would you feel about this?

Not satisfied at all	Mostly dissatisfied	Somewhat satisfied	Mostly satisfied	Highly satisfied
❑	❑	❑	❑	❑

ii. How often do you worry that you may have a heart attack or die suddenly?

I can't stop worrying about it	**I often** think or worry about it	**I occasionally** worry about it	**I rarely** think or worry about it	**I never** think or worry about it
❑	❑	❑	❑	❑

References

1. Berwick DM. Medical associations: guilds or leaders? BMJ. 1997;314(7094):1564-5.
2. Devlin N. Getting the most out of PROMs. The Kings Fund; 2010.
3. Statistics OoN. Deaths registered in England 2013 ONS2019 [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsregistrationsummarytables/2014-07-16.
4. Cutler D Me. Changes in the Age Distribution of Mortality Over the 20th Century: The National Bureau of Economic Research; 2001 [NBER Working Paper No. 8556:[Available from: https://www.nber.org/papers/w8556.
5. Dept_of_Health. Stoptober starts today: Department of Health; 2012 [Available from: https://webarchive.nationalarchives.gov.uk/20160117231857/https://www.gov.uk/government/news/stoptober-starts-today
6. Barham LD, N. Patient-reported outcome measures: Implications for nursing. Nursing Standard. 2011;25(18):42-5.
7. Darzi. NHS Next Stage Review Interim Report. Dept of Health; 2007.
8. Mackintosh AC, C; Fitzpatrick, R. A Structured Review of Patient Reported Outcome Measures Used In Elective Procedures for Coronary Revascularisation Department of Health; 2010.
9. Penson DF. The effect of erectile dysfunction on quality of life following treatment for localized prostate cancer. Rev Urol. 2001;3(3):113-9.
10. Chang EM, Gillespie EF, Shaverdian N. Truthfulness in patient-reported outcomes: factors affecting patients' responses and impact on data quality. Patient Relat Outcome Meas. 2019;10:171-86.
11. Helgason AR, Adolfsson J, Dickman P, Fredrikson M, Arver S, Steineck G. Waning sexual function--the most important disease-specific distress for patients with prostate cancer. Br J Cancer. 1996;73(11):1417-21.
12. Studies IoF. The IFS Green Budget February 2017. 2017 Feb 2017.
13. Roth GA, Johnson C, Abajobir A, Abd-Allah F, Abera SF, Abyu G, et al. Global, Regional, and National Burden of Cardiovascular Diseases for 10 Causes, 1990 to 2015. J Am Coll Cardiol. 2017;70(1):1-25.
14. Bhatnagar P, Wickramasinghe K, Wilkins E, Townsend N. Trends in the epidemiology of cardiovascular disease in the UK. Heart. 2016;102(24):1945-52.
15. Luengo-Fernandez R, Leal J, Gray A, Petersen S, Rayner M. Cost of cardiovascular diseases in the United Kingdom. Heart. 2006;92(10):1384-9.
16. Gershlick T, Thomas M. PCI or CABG: which patients and at what cost? Heart. 2007;93(10):1188-90.
17. Weintraub WS, Spertus JA, Kolm P, Maron DJ, Zhang Z, Jurkovitz C, et al. Effect of PCI on quality of life in patients with stable coronary disease. N Engl J Med. 2008;359(7):677-87.
18. Serruys PW, Ong AT, van Herwerden LA, Sousa JE, Jatene A, Bonnier JJ, et al. Five-year outcomes after coronary stenting versus bypass surgery for the treatment of multivessel disease: the final analysis of the Arterial Revascularization Therapies Study (ARTS) randomized trial. J Am Coll Cardiol. 2005;46(4):575-81.
19. Cohen DJ, Osnabrugge RL, Magnuson EA, Wang K, Li H, Chinnakondepalli K, et al. Cost-effectiveness of percutaneous coronary intervention with drug-eluting stents versus bypass surgery for patients with 3-vessel or left main coronary artery disease: final results from the Synergy Between Percutaneous Coronary Intervention With TAXUS and Cardiac Surgery (SYNTAX) trial. Circulation. 2014;130(14):1146-57.

20. Black N. Patient reported outcome measures could help transform healthcare. BMJ. 2013;346:f167.
21. Griggs CL, Schneider JC, Kazis LE, Ryan CM. Patient-reported Outcome Measures: A Stethoscope for the Patient History. Ann Surg. 2017;265(6):1066-7.
22. Bren L. The Importance of Patient Reported Outcomes Its al about the Patients [Web page]. US Food and Drug Administration; 2006 [Available from: https://permanent.access.gpo.gov/lps1609/www.fda.gov/fdac/features/2006/606_patients.html.
23. Douglas SL, Daly BJ, Meropol NJ, Lipson AR. Patient-physician discordance in goals of care for patients with advanced cancer. Curr Oncol. 2019;26(6):370-9.
24. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276-82.
25. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med. 2005;37(5):360-3.
26. Dept_of_Health. Guidance on the Routine Collection of Patient Reported Outcome Measures (PROMs). 2008.
27. Dept_of_Health. High Quality Care For All. Department of Health; 2010.
28. EuroQol. EQ-5D-5L User Guide 2015 [User Guide]. Available from: www.euroqol.org.
29. Maruish M. Users Manual For The SF-36 v2 Health Survey. User Manual. Medical Outcomes Trust; 2011 2011.
30. Spertus JA, Winder JA, Dewhurst TA, Deyo RA, Prodzinski J, McDonell M, et al. Development and evaluation of the Seattle Angina Questionnaire: a new functional status measure for coronary artery disease. J Am Coll Cardiol. 1995;25(2):333-41.
31. Bass M, Morris S, Neapolitan R. Utilizing Multidimensional Computer Adaptive Testing to Mitigate Burden With Patient Reported Outcomes. AMIA Annu Symp Proc. 2015;2015:320-8.
32. Vodicka E, Kim K, Devine EB, Gnanasakthy A, Scoggins JF, Patrick DL. Inclusion of patient-reported outcome measures in registered clinical trials: Evidence from ClinicalTrials.gov (2007-2013). Contemp Clin Trials. 2015;43:1-9.
33. SoS_Investigators. Coronary artery bypass surgery versus percutaneous coronary intervention with stent implantation in patients with multivessel coronary artery disease (the Stent or Surgery trial): a randomised controlled trial. Lancet. 2002;360(9338):965-70.
34. Serruys PW, Unger F, Sousa JE, Jatene A, Bonnier HJ, Schonberger JP, et al. Comparison of coronary-artery bypass surgery and stenting for the treatment of multivessel disease. N Engl J Med. 2001;344(15):1117-24.
35. King SB, 3rd, Lembo NJ, Weintraub WS, Kosinski AS, Barnhart HX, Kutner MH, et al. A randomized trial comparing coronary angioplasty with coronary bypass surgery. Emory Angioplasty versus Surgery Trial (EAST). N Engl J Med. 1994;331(16):1044-50.
36. Weintraub WS, Mauldin PD, Becker E, Kosinski AS, King SB, 3rd. A comparison of the costs of and quality of life after coronary angioplasty or coronary surgery for multivessel coronary artery disease. Results from the Emory Angioplasty Versus Surgery Trial (EAST). Circulation. 1995;92(10):2831-40.
37. Hampton JR. Coronary angioplasty versus coronary artery bypass surgery: the Randomized Intervention Treatment of Angina (RITA) trial. Lancet. 1993;341(8845):573-80.
38. Bypass Angioplasty Revascularization Investigation I. Comparison of coronary bypass surgery with angioplasty in patients with multivessel disease. N Engl J Med. 1996;335(4):217-25.

39. Ware JS, Kristin; Kosinski, Mark; Gandek, Barbara. SF-36 Health Survey Manual & Interpretation Guide: The Health Institute, New England Medical Centre; 1993. 316 p.
40. Campeau L. Letter: Grading of angina pectoris. Circulation. 1976;54(3):522-3.
41. Foundation; ER. EQ-5D-5L User Guide Online: EuroQol Group; 2017 [updated 18/04/2017. Basic Information on how to use the EQ-5D-L instrument]. Available from: https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/.
42. Wahrborg P, Emanuelsson H. The cardiac health profile: content, reliability and validity of a new disease-specific quality of life questionnaire. Coron Artery Dis. 1996;7(11):823-9.
43. Zhang Z, Mahoney EM, Stables RH, Booth J, Nugara F, Spertus JA, et al. Disease-specific health status after stent-assisted percutaneous coronary intervention and coronary artery bypass surgery: one-year results from the Stent or Surgery trial. Circulation. 2003;108(14):1694-700.
44. Wahrborg P, Booth JE, Clayton T, Nugara F, Pepper J, Weintraub WS, et al. Neuropsychological outcome after percutaneous coronary intervention or coronary artery bypass grafting: results from the Stent or Surgery (SoS) Trial. Circulation. 2004;110(22):3411-7.
45. Zhang Z, Mahoney EM, Spertus JA, Booth J, Nugara F, Kolm P, et al. The impact of age on outcomes after coronary artery bypass surgery versus stent-assisted percutaneous coronary intervention: one-year results from the Stent or Surgery (SoS) trial. Am Heart J. 2006;152(6):1153-60.
46. Zhang Z, Spertus JA, Mahoney EM, Booth J, Nugara F, Stables RH, et al. The impact of acute coronary syndrome on clinical, economic, and cardiac-specific health status after coronary artery bypass surgery versus stent-assisted percutaneous coronary intervention: 1-year results from the stent or surgery (SoS) trial. Am Heart J. 2005;150(1):175-81.
47. Anker SD, Agewall S, Borggrefe M, Calvert M, Jaime Caro J, Cowie MR, et al. The importance of patient-reported outcomes: a call for their comprehensive integration in cardiovascular clinical trials. Eur Heart J. 2014;35(30):2001-9.
48. WHO. WHOQOL: Measuring Quality of Life Online2020 [Available from: https://www.who.int/healthinfo/survey/whoqol-qualityoflife/en/.
49. Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities. Br Med Bull. 2010;96:5-21.
50. Weinstein MC, Torrance G, McGuire A. QALYs: the basics. Value Health. 2009;12 Suppl 1:S5-9.
51. Hamilton DF, Lane JV, Gaston P, Patton JT, Macdonald D, Simpson AH, et al. What determines patient satisfaction with surgery? A prospective cohort study of 4709 patients following total joint replacement. BMJ Open. 2013;3(4).
52. Fournier SM, DG. Rediscovering Satisfaction. Journal of Marketing. 1999;63(4):18.
53. Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. BMJ Open. 2013;3(1).
54. Lim CR, Harris K, Dawson J, Beard DJ, Fitzpatrick R, Price AJ. Floor and ceiling effects in the OHS: an analysis of the NHS PROMs data set. BMJ Open. 2015;5(7):e007765.
55. Thompson P. What is a QALY online: Haywood Medical; 2001 [updated May 2001. Vol 1 number 6:[Available from: www.evidence-based-medicine.co.uk.
56. Drozda J, Jr., Messer JV, Spertus J, Abramowitz B, Alexander K, Beam CT, et al. ACCF/AHA/AMA-PCPI 2011 performance measures for adults with coronary artery disease and hypertension: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Performance Measures and the

American Medical Association-Physician Consortium for Performance Improvement. Circulation. 2011;124(2):248-70.
57. Chan PS, Jones PG, Arnold SA, Spertus JA. Development and validation of a short version of the Seattle angina questionnaire. Circ Cardiovasc Qual Outcomes. 2014;7(5):640-7.
58. Stewart A L HRD, Ware JE. Measuring functioning and well-being: The Medical Outcomes Study approach. Durham, NC: Duke University Press; 1992.
59. Rand_Health_Care. 36 Item Short Form Survey (SF-36) Scoring Instructions Online2020 [Available from: www.rand.org/heslth-care/survey_tools/mos/36-item-short-form/scoring.html.
60. Andresen EM, Gravitt GW, Aydelotte ME, Podgorski CA. Limitations of the SF-36 in a sample of nursing home residents. Age Ageing. 1999;28(6):562-6.
61. Iliceto A, Berndt SL, Greenslade JH, Parsonage WA, Hammett C, Than M, et al. Agreement Between Patient-reported and Cardiology-adjudicated Medical History in Patients With Possible Ischemic Chest Pain: An Observational Study. Crit Pathw Cardiol. 2016;15(3):121-5.
62. Borkon AM, Muehlebach GF, House J, Marso SP, Spertus JA. A comparison of the recovery of health status after percutaneous coronary intervention and coronary artery bypass. Ann Thorac Surg. 2002;74(5):1526-30; discussion 30.
63. Kemp I, Appleby C, Lane S, Lisboa P, Stables RH. A comparison of angina symptoms reported by clinicians and patients, pre and post revascularisation: Insights from the Stent or Surgery Trial. Int J Cardiol. 2019;293:25-31.
64. Pakhomov SV, Jacobsen SJ, Chute CG, Roger VL. Agreement between patient-reported symptoms and their documentation in the medical record. Am J Manag Care. 2008;14(8):530-9.
65. Elfil M, Negida A. Sampling methods in Clinical Research; an Educational Review. Emerg (Tehran). 2017;5(1):e52.
66. Shafiq A, Arnold SV, Gosch K, Kureshi F, Breeding T, Jones PG, et al. Patient and physician discordance in reporting symptoms of angina among stable coronary artery disease patients: Insights from the Angina Prevalence and Provider Evaluation of Angina Relief (APPEAR) study. Am Heart J. 2016;175:94-100.
67. Rumsfeld JS, Alexander KP, Goff DC, Jr., Graham MM, Ho PM, Masoudi FA, et al. Cardiovascular health: the importance of measuring patient-reported health status: a scientific statement from the American Heart Association. Circulation. 2013;127(22):2233-49.
68. Sewitch MJ, Abrahamowicz M, Dobkin PL, Tamblyn R. Measuring differences between patients' and physicians' health perceptions: the patient-physician discordance scale. J Behav Med. 2003;26(3):245-64.
69. Berkanovic E, Hurwicz ML, Lachenbruch PA. Concordant and discrepant views of patients' physical functioning. Arthritis Care Res. 1995;8(2):94-101.
70. Engel GL. The need for a new medical model: a challenge for biomedicine. Science. 1977;196(4286):129-36.
71. Barton JL, Imboden J, Graf J, Glidden D, Yelin EH, Schillinger D. Patient-physician discordance in assessments of global disease severity in rheumatoid arthritis. Arthritis Care Res (Hoboken). 2010;62(6):857-64.
72. Desthieux CM, A.; Granger, B.; Saraux, A.; Fautrel, B.; Gossec, L. Patient-physician discordance in global assessment in early spondyloarthritis and its change over time: the DESIR cohort. Annals of the Rheumatic Diseases. 2015;75:5.
73. Dobkin PL, De Civita M, Abrahamowicz M, Bernatsky S, Schulz J, Sewitch M, et al. Patient-physician discordance in fibromyalgia. J Rheumatol. 2003;30(6):1326-34.

74. Henderson A, Andreyev HJ, Stephens R, Dearnaley D. Patient and physician reporting of symptoms and health-related quality of life in trials of treatment for early prostate cancer: considerations for future studies. Clin Oncol (R Coll Radiol). 2006;18(10):735-43.
75. Di Maio M, Gallo C, Leighl NB, Piccirillo MC, Daniele G, Nuzzo F, et al. Symptomatic toxicities experienced during anticancer treatment: agreement between patient and physician reporting in three randomized trials. J Clin Oncol. 2015;33(8):910-5.
76. Thomas SH, Borczuk P, Shackelford J, Ostrander J, Silver D, Evans M, et al. Patient and physician agreement on abdominal pain severity and need for opioid analgesia. Am J Emerg Med. 1999;17(6):586-90.
77. Tory H, Zurakowski D, Kim S, Investigators CJDQMWftCR. Patient and physician discordance of global disease assessment in juvenile dermatomyositis: findings from the Childhood Arthritis & Rheumatology Research Alliance Legacy Registry. Pediatr Rheumatol Online J. 2020;18(1):5.
78. Jacome C, Pereira AM, Almeida R, Ferreira-Magalhaes M, Couto M, Araujo L, et al. Patient-physician discordance in assessment of adherence to inhaled controller medication: a cross-sectional analysis of two cohorts. BMJ Open. 2019;9(11):e031732.
79. Desthieux C, Hermet A, Granger B, Fautrel B, Gossec L. Patient-Physician Discordance in Global Assessment in Rheumatoid Arthritis: A Systematic Literature Review With Meta-Analysis. Arthritis Care Res (Hoboken). 2016;68(12):1767-73.
80. Sewitch MJ, Abrahamowicz M, Bitton A, Daly D, Wild GE, Cohen A, et al. Psychosocial correlates of patient-physician discordance in inflammatory bowel disease. Am J Gastroenterol. 2002;97(9):2174-83.
81. Ghukasyan L, Etcheto A, Dougados M, Molto A. Patient and physician agreement on reported Bath Ankylosing Spondylitis Disease Activity Index in patients with axial spondyloarthritis. Joint Bone Spine. 2017;84(6):749-50.
82. ND F-RLAPALRRPD-PCBBBPM. When patients report diseases that prescribers seem unaware of: Discordance between patient and physician reporting of risk related previous history in NSAID users from the CADEUS study. Clinical Pharmacology and Therapeutics. 2010;88(Nov 2010):8.
83. Tago M, Sawada T, Nishiyama S, Tahara K, Kato E, Hayashi H, et al. Influence of large joint involvement on patient-physician discordance in global assessment of rheumatoid arthritis disease activity analyzed by a novel joint index. Int J Rheum Dis. 2018;21(6):1237-45.
84. Wang CTM, Kwan YH, Fong W, Xiong SQ, Leung YY. Factors associated with patient-physician discordance in a prospective cohort of patients with psoriatic arthritis: An Asian perspective. Int J Rheum Dis. 2019;22(7):1209-15.
85. Wang CTM, Fong W, Kwan YH, Phang JK, Lui NL, Leung YY, et al. A cross-sectional study on factors associated with patient-physician discordance in global assessment of patients with axial spondyloarthritis: an Asian perspective. Int J Rheum Dis. 2018;21(7):1436-42.
86. Desthieux C, Granger B, Balanescu AR, Balint P, Braun J, Canete JD, et al. Determinants of Patient-Physician Discordance in Global Assessment in Psoriatic Arthritis: A Multicenter European Study. Arthritis Care Res (Hoboken). 2017;69(10):1606-11.
87. Sonnenberg. Personal View: patient-physician discordance about benefits and risks in gastroenterology decision-making. Alimentary Pharmacology and Therapeutics. 2004;19:6.
88. Beltrame JF, Weekes AJ, Morgan C, Tavella R, Spertus JA. The prevalence of weekly angina among patients with chronic stable angina in primary care practices: The

Coronary Artery Disease in General Practice (CADENCE) Study. Arch Intern Med. 2009;169(16):1491-9.
89. Stables RH. Design of the 'Stent or Surgery' trial (SoS): a randomized controlled trial to compare coronary artery bypass grafting with percutaneous transluminal coronary angioplasty and primary stent implantation in patients with multi-vessel coronary artery disease. Semin Interv Cardiol. 1999;4(4):201-7.
90. SoS Investigators. Coronary artery bypass surgery versus percutaneous coronary intervention with stent implantation in patients with multivessel coronary artery disease (the Stent or Surgery trial): a randomised controlled trial. Lancet. 2002;360(9338):965-70.
91. Lowly R. Kappa statistic 2018 [Available from: http://vassarstats.net/kappa.html.
92. Campeau L. The Canadian Cardiovascular Society grading of angina pectoris revisited 30 years later. Can J Cardiol. 2002;18(4):371-9.
93. Stone GW, Ellis SG, Gori T, Metzger DC, Stein B, Erickson M, et al. Blinded outcomes and angina assessment of coronary bioresorbable scaffolds: 30-day and 1-year results from the ABSORB IV randomised trial. Lancet. 2018;392(10157):1530-40.
94. Gluud LL. Bias in clinical intervention research. Am J Epidemiol. 2006;163(6):493-501.
95. Chapman EN, Kaatz A, Carnes M. Physicians and implicit bias: how doctors may unwittingly perpetuate health care disparities. J Gen Intern Med. 2013;28(11):1504-10.
96. Fischer ME, Gill. When doctors and patients talk: making sense of the consultation. Health Foundation

2012 2012.
97. Cox JL, Naylor CD, Johnstone DE. Limitations of Canadian Cardiovascular Society classification of angina pectoris. Am J Cardiol. 1994;74(3):276-7.
98. Cox J, Naylor CD. The Canadian Cardiovascular Society grading scale for angina pectoris: is it time for refinements? Ann Intern Med. 1992;117(8):677-83.
99. Rombach I, Gray AM, Jenkinson C, Murray DW, Rivero-Arias O. Multiple imputation for patient reported outcome measures in randomised controlled trials: advantages and disadvantages of imputing at the item, subscale or composite score level. BMC Med Res Methodol. 2018;18(1):87.
100. Simons CL, Rivero-Arias O, Yu LM, Simon J. Multiple imputation to deal with missing EQ-5D-3L data: Should we impute individual domains or the actual index? Qual Life Res. 2015;24(4):805-15.
101. Kimble LP, Dunbar SB, Weintraub WS, McGuire DB, Fazio S, De AK, et al. The Seattle angina questionnaire: reliability and validity in women with chronic stable angina. Heart Dis. 2002;4(4):206-11.
102. Vizintin Z. Why five and not eight? How round number bias can reduce your nest egg 2018 [Available from: behavioralscientist.org/five-not-eight-round-number-bias-can-reduce-nest-egg/.
103. Jain GG, Gary; Nayakankuppam, D; Levin, I. Revisiting attribute framing: The impact of number roundedness on framing. Organisational Behaviour and Human Decision Processes. 2020;161:10.
104. Giavarina D. Understanding Bland Altman analysis. Biochem Med (Zagreb). 2015;25(2):141-51.
105. Mercieca-Bebber R, Calvert M, Kyte D, Stockler M, King MT. The administration of patient-reported outcome questionnaires in cancer trials: Interviews with trial coordinators regarding their roles, experiences, challenges and training. Contemp Clin Trials Commun. 2018;9:23-32.
106. Bowling A. Mode of questionnaire administration can have serious effects on data quality. J Public Health (Oxf). 2005;27(3):281-91.

107. Tourangeau RR, L; Rasinski, K. 10: Mode of Data Collection: Cambridge University Press; 2000.
108. Spertus J. Barriers to the use of patient-reported outcomes in clinical care. Circ Cardiovasc Qual Outcomes. 2014;7(1):2-4.
109. Garratt AM, Hutchinson A, Russell I, Network for Evidence-Based Practice in N, Yorkshire. The UK version of the Seattle Angina Questionnaire (SAQ-UK): reliability, validity and responsiveness. J Clin Epidemiol. 2001;54(9):907-15.
110. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philos Trans A Math Phys Eng Sci. 2016;374(2065):20150202.
111. Arnold SV, Kosiborod M, Li Y, Jones PG, Yue P, Belardinelli L, et al. Comparison of the Seattle Angina Questionnaire With Daily Angina Diary in the TERISA Clinical Trial. Circ Cardiovasc Qual Outcomes. 2014;7(6):844-50.
112. Glen S. Recall Bias: Definition, Examples, Strategies to Avoid it [Web paqge]. 2020 [updated 27/05/2016. Available from: https://www.statisticshowto.com/recall-bias/.
113. Dougherty CM, Dewhurst T, Nichol WP, Spertus J. Comparison of three quality of life instruments in stable angina pectoris: Seattle Angina Questionnaire, Short Form Health Survey (SF-36), and Quality of Life Index-Cardiac Version III. J Clin Epidemiol. 1998;51(7):569-75.
114. Spertus JA. Evolving applications for patient-centered health status measures. Circulation. 2008;118(20):2103-10.
115. Borgia F, Viceconte N, Ali O, Stuart-Buttle C, Saraswathyamma A, Parisi R, et al. Improved cardiac survival, freedom from MACE and angina-related quality of life after successful percutaneous recanalization of coronary artery chronic total occlusions. Int J Cardiol. 2012;161(1):31-8.
116. Arnold SV, Spertus JA, Jones PG, Xiao L, Cohen DJ. The impact of dyspnea on health-related quality of life in patients with coronary artery disease: results from the PREMIER registry. Am Heart J. 2009;157(6):1042-9 e1.
117. Almalki S. Integrating Quantitative and Qualitative Data in Mixed Methods Research: Challenges and Benefits. Journal of Education and Learning. 2016;5(3):9.
118. Rovai AB, JD; Ponton, MK. Social Science Research Design and Statistics. Chesapeake, VA: Watertree Press LLC; 2014.
119. LeCompte M. The transformation of ethnographic practice: Past and current challenges: Sage, London; 2002.
120. Lelkes YK, JA; Marx, DM; Judd, CM; Park, B. Complete anonymity compromises the accuracy of self-reports. J of Experimental Social Psychology. 2012;48(6):8.
121. O'Reilly-Shah VN. Factors influencing healthcare provider respondent fatigue answering a globally administered in-app survey. PeerJ. 2017;5:e3785.
122. De Schrijver A. Sample Survey on Sensitive Topics: Investigating Respondents' Understanding and Trust in Alternative Versions of the Randomized Response Technique. Journal of Research Practice. 2012;8(1).
123. Block D. Exploring interpretations of questionnaire items. System; 1998.
124. Choi BP, A. A catalogue of biases in questionnaires. Preventing Chronic Disease. 2005;2(1):13.
125. Jovancic N. Types of Bias in Research and How to Make Your Surveys Bias Free Online2019 [Available from: https://www.leadquizzes.com/blog/types-of-bias-in-research/.
126. Mahtani KS, EA; Brassey, J. Observer Bias Web page: University of Oxford; 2017 [Available from: https://www.catalogofbias.org/biases/observer-bias.
127. Society MOfSa. Would Osler stand by his famous quote today? : McGill Office for Science and Society; 2020 [Available from:

https://www.mcgill.ca/oss/article/controversial-science-health-history-news/would-osler-stand-his-famous-quote-today.
128. Cooper J. Healthcare expenditure, UK Health Accounts: 2018 Online: Office of National Statistics; 2020 [updated 28/04/2020. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthcaresystem/bulletins/ukhealthaccounts/2018#:~:text=1.-,Main%20points,2017%20and%206.9%25%20in%201997.