# MASTER OF PHILOSOPHY

## Exploring the impact of using different researchers to capture Quality of Life measures in Dementia randomised controlled trials

Evans, Rachel

*Award date:*
2020

*Awarding institution:*
Bangor University

# Exploring the impact of using different researchers to capture Quality of Life measures in Dementia randomised controlled trials


PRIFYSGOL
BANGOR
UNIVERSITY

Rachel Evans

Student ID: 500303742

March 2020

Thesis Presented for MPhil

College of Human Sciences, School of Health Sciences, Bangor University

**Acknowledgements**

I would like to thank both my supervisors Paul Brocklehurst and Zoë Hoare for guiding and mentoring me through this project. I would additionally like to thank Jean Ryan for her input and additional support. Without their encouragement, advice and help I would not have completed this thesis. They kept me motivated and engaged, pushing me towards deadlines (which I needed) whilst being extremely understanding of other work commitments and personal circumstances. This not only applies to this work, but I also attribute my progression at the trials unit to their ability to see my potential and will be forever grateful for this.

I would also like to thank the rest of the NWORTH team who are all very supportive of this project and have been a delight to work with which has made this journey that bit easier. I would particularly like to thank two previous NWORTH team members Nikki Totton and Lisa Jones who were also very supportive, encouraging and helpful - whilst they were colleagues and since leaving NWORTH.

I am exceptionally grateful to my parents, partner, sister and friends for the help and support I have received in completing this project and throughout many other adventures in my life. Their continuous backing and support has been vital in everything I have achieved.

Finally, this thesis would not be possible without the use of the data from the REMCARE study and I am extremely thankful to Robert Woods, the Chief Investigator of the study, for allowing the use of the data.

# Contents

**Abbreviations**

| | |
|---|---|
| AD | Alzheimer's disease |
| ADRQL | Alzheimer's Disease Related Quality of Life |
| ANCOVA | Analysis of Covariance |
| BASQID | Bath Assessment of Subjective Quality of Life in Dementia |
| BMJ | British Medical Journal |
| CBD | Cannabidiol |
| CTU | Clinical Trials Unit |
| DEMQOL | Dementia Quality of Life questionnaire |
| DQOL | Dementia Quality of life Questionnaire |
| DSM-IV | Diagnostic and Statistical Manual of Mental Disorders |
| EBM | Evidence Based Medicine |
| HRQoL | Health Related Quality of Life |
| ICC | intraclass correlation coefficient |
| iCST | the individual Cognitive Stimulation Therapy |
| JtD | Journeying through Dementia |
| LMM | Linear mixed model |
| NHS | National Health Service |
| NWORTH | North Wales Organisation for Randomised Trials in Health |
| PES-AD | Pleasant Events Schedule in Alzheimer's Disease |
| POM | primary outcome measure |
| PROMs | patient reported outcome measures |
| PwD | People with Dementia |
| QCPR | Quality of the Caregiving Relationship |
| QoL | Quality of Life |
| QOL-AD | Quality of Life in Alzheimer's Disease |
| QOLAS | Quality of Life Assessment Schedule |
| QUALID | Quality of life in late-stage Dementia |
| RCT | Randomised Controlled Trial |
| REMCARE | Reminiscence groups for people with dementia and their family caregivers |
| SWAT | Study within a Trial |
| WHELD | Well-Being and Health for People with Dementia |
| WHO | World Health Organization |
| UK | United Kingdom |

## Thesis outline

This thesis examines the impact of using different researchers to capture quality of life outcome measures in Dementia randomised controlled trials. Chapter One starts with an introduction to Randomised Controlled Trials (RCT) and their role in evidence-based medicine and where they sit in the evidence hierarchy, before exploring the different types of biases that can arise in RCTs, which can impact upon research quality. The focus then turns to the importance of outcome measurement in RCTs and explores the issues of measurement error and the specific sources of instrument bias, particularly the impact of researcher bias on outcome measurement.

Chapter Two provides a brief introduction to Dementia and the impact Dementia can have on the quality of life of patients and their carers. The focus then turns to measuring quality of life in Dementia RCTs and the specific issues associated with this. Chapter Two concludes by focusing on the issues of instrument bias in Dementia RCTs and the influence of using multiple researchers to gather outcome data, before outlining the research questions to be addressed in this thesis.

Chapters Three describes the data used for the statistical analysis and includes a brief introduction to the RCT from which the data was derived and extracted. Chapters Four and Five detail the results of the statistical analysis conducted in order to assess the two different research questions proposed.

The thesis concludes in Chapter Six with a discussion of the results, a comparison of these results to other literature and the detail of the statistical methods adopted in comparison to other approaches used. It also describes the main limitations of the thesis and includes recommendations for future research and implications for current practice. The chapter ends with the final conclusions from the thesis, based on the analysis undertaken and its interpretation.

# Chapter One – Introduction & Background

## 1.1 The randomised controlled trial

Randomised Controlled Trials (RCTs) have been central to the Evidence Based Medicine (EBM) paradigm and research for many decades (Bothwell & Podolsky, 2016; Pawson, 2013). The first recognised RCT was published in 1948 by the British Medical Journal (BMJ) (Bothwell & Podolsky, 2016; Crofton & Mitchison, 1948; McDonald et al., 2002). This study assessed the effects of streptomycin in patients with tuberculosis, although quasi-experimental methodologies were in use long before this (Bothwell & Podolsky, 2016; Crofton & Mitchison, 1948; Lindsay, 2004; McDonald et al., 2002). RCTs adopt a positivist approach to the evaluation of interventions: "the idea is to create two identical systems into one of which a new component [the intervention] is introduced" (p. 4) (Pawson, 2013). Using an experimental paradigm "observations are then made of outcome differences that occur between experimental and control conditions and should a change occur, it is attributed to the one difference between them" (p. 4) (Pawson, 2013). Other factors that could influence the outcome (known as 'confounders') are either taken account *a priori* (e.g. by the stratification of baseline variables) or adjusted for *post hoc* (e.g. using multiple regression). RCTs are based on the underlying assumption that having controlled for other variables, only the intervention produces the observed effect. As such, RCTs are the only study design that enable researchers to establish causality rather than association.

EBM encourages health care professionals to use the best quality evidence in the decisions that inform their clinical practice. The evidence hierarchy is a framework used to rank research methodologies according to the rigor of their design (Akobeng, 2005). Commonly, the hierarchy is presented pictorially as a pyramid that consists of a series of levels. The higher up the pyramid, the more rigorous the research design is purported to be (Akobeng, 2005). As RCTs are able to demonstrate causality, they are considered to be 'the Gold standard' of primary research and sit at the top of the evidence hierarchy (Murad et al., 2016; Siepmann et al., 2016), only surpassed by systematic reviews of RCTs (secondary research). However, this form of categorisation only focuses on the type of study design undertaken, with little consideration for the **quality** of the underpinning research. It is also founded on the philosophical principle of empiricism i.e. that we can measure important phenomena and meaningful change over time. This assumes that the measures that we use are stable and are not affected by how they are used by the researcher. Testing this assumption forms the underlying basis for this thesis.

Modifications to the traditional pyramid have recently been proposed expanding the categorisation process to focus on the issue of study quality (Murad et al., 2016). This highlights the fact that the conduct of the research is as important as the underlying design of the study. As such, Murad et al. (2016) propose that the lines between the different levels in the evidence hierarchy should be wavy rather than straight (Figure 1). For example, they argue that a well conducted Cohort study should be regarded as highly, or higher, than a poorly conducted RCT, as the conduct of the study counter-balances the rigor of the design.

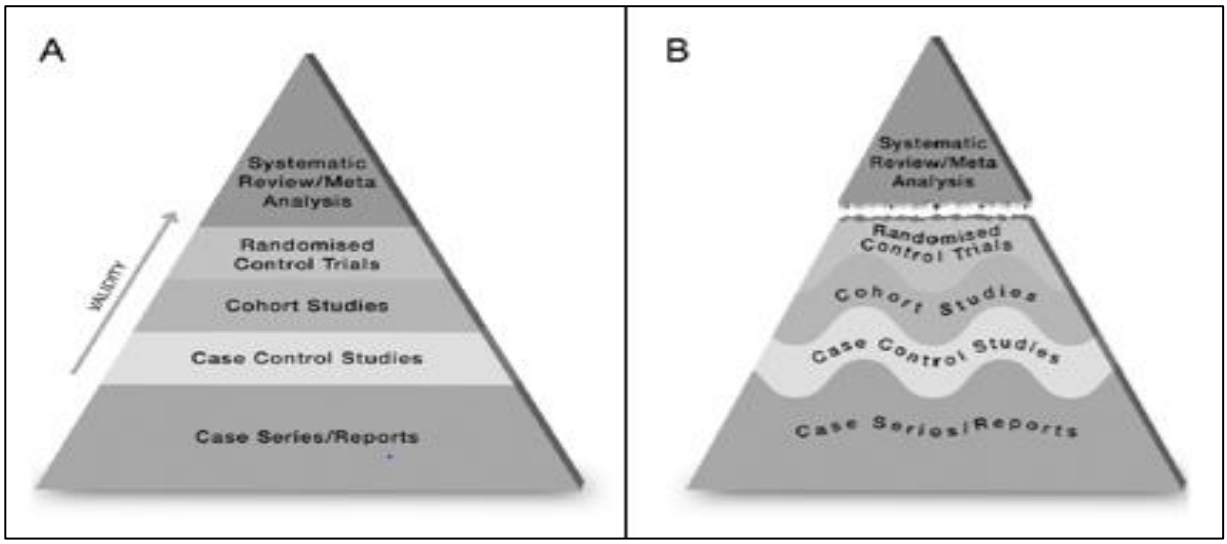


***Figure 1**: (A) Traditional Evidence Hierarchy Pyramid (B) Suggested Amendments to Pyramid.* Source: (*Murad et al., 2016*)

## 1.2 Bias in Randomised Controlled Trials

Typically, external and internal validity are used as indications of RCT quality (Akobeng, 2005). External validity relates to the generalisability of the trial findings, i.e. how representative they are of the context that the intervention will eventually be embedded within (Juni, 2001; Rothwell, 2005). Internal validity refers to the quality and rigor of the trial design (other than sample size) and its conduct (Gluud, 2006; Higgins et al., 2011; Juni, 2001). As such, external validity is highly dependent on the design of the intervention and the type of participants recruited, whilst internal validity relies on the rigor applied to the research process and reducing study bias (Gluud, 2006; Higgins et al., 2011; Siepmann et al., 2016).

Although RCTs are seen to sit at the top of the primary evidence hierarchy, they can be complex to deliver and subject to different types of study bias (Pandis, 2011). Broadly, bias can be classed into two forms: random bias (i.e. random error) and systematic bias (Malone et al., 2014). The

former relates to the notion that regardless of how well a study is designed and conducted, there is always a statistical chance that some form of random error can occur within the complexity of study delivery (Pandis, 2011). In an inferential model, Type I (stating an effect is present when it is not) and Type II errors (stating an effect isn't there when it is) are reduced by careful attention to the sample size calculation, where both types of error are reduced as much as possible (Siepmann et al., 2016).

Systematic error is classed as an error that arises during the design and conduct of an RCT, i.e. it does not occur randomly, but is caused by elements within the design of the trial or due to the actions of the study team, which are often repeated during the course of the study (Malone et al., 2014; Tripepi et al., 2008). From a statistical perspective, systematic error is defined as 'a systematic distortion of a statistical result due to a factor not allowed for in its derivation' ("A Word About Evidence: 6. Bias — a proposed definition", n.d.). The key difference between random and systematic error is that the former is a 'one-off' event, whereas the latter is repeated throughout the study, which then distorts the magnitude and direction of the effect seen and threatens the quality of the research undertaken.

There is potential to introduce bias at any stage of the research process (and one RCT may have multiple sources of bias in play at any one point in time (Gluud, 2006; Malone et al., 2014; Pandis, 2011; Siddiqi, 2011; Smith & Noble, 2014). Furthermore, the introduction of bias into an RCT may not necessarily be intentional (Simundic, 2013). Whilst the deliberate distortion of RCT processes is unethical, unintentional bias has a similar impact on the research process and should be treated with the same caution (Simundic, 2013). A common misconception is that an RCT either contains bias or does not, when in reality, the presence of bias is not a binary concept (Gerhard, 2008; Pannucci & Wilkins, 2010). No research study is completely free of bias nor is it completely biased. Instead, the focus of study quality should be around ***how much*** bias is present and to what extent bias impacts upon the conclusions drawn from the trial (Gerhard, 2008; Pannucci & Wilkins, 2010).

Collectively, the presence of bias impacts on the internal validity of an RCT (Malone et al., 2014; Smith & Noble, 2014). Systematic and repetitive distortions can lead to an over or under estimation of the treatment effect, thereby creating uncertainty and making results and conclusions difficult to interpret (Malone et al., 2014; Marshall et al., 2015). In turn, this directly impacts upon policy maker's decisions. If commissioners and clinicians implement new procedures and processes based on the findings of an RCT with a substantive level of bias, this could at the very least produce no health effect, or at its worst, cause harm to patients (Gerhard, 2008; Simundic, 2013; Smith &

Noble, 2014). Although bias cannot be prevented entirely and is often difficult to omit, many techniques adopted by trialists aim to reduce bias. Equally, given the damaging effect of drawing an incorrect inference, there is an ethical imperative for all investigators to act to minimise the level of bias in RCTs and ensure that the whole research team are aware of potential biases and their impact (Malone et al., 2014; Simundic, 2013; Smith & Noble, 2014).

Whilst there are many types of biases possible in RCTs, five main domains have been identified by the Cochrane Bias Methods Group: selection bias, performance bias, detection bias, attrition bias and reporting bias (Higgins et al., 2017). Although Cochrane focus on these five sources of bias, there are many other types of biases that can emerge in RCTs such as publication bias (Allen, 2002; Siddiqi, 2011), analysis bias (Indrayan, 2012; Simundic, 2013), contamination bias (Arnold, 2011; Indrayan, 2012), compliance bias and many more that go beyond the scope of this thesis. The Table in Appendix 1 summarises these main forms of bias and their impacts on RCTs, but detail is provided below on the fundamental processes within RCTs that help to reduce bias.

One of the fundamental assumptions underlying RCTs is that researchers have '*clinical equipoise'* i.e. none of the interventions to be tested are *known* to be superior to another, *a priori* (Cook & Sheets, 2011). For this to hold, there must be genuine uncertainty about the effects of the interventions being tested (Cook & Sheets, 2011). The assumption is considered to be a necessary condition for RCTs as it allows the application of some of the techniques adopted in RCTs to be conducted whilst minimising bias (Hey et al., 2017; Siepmann et al., 2016). In order to maintain equipoise, ideally, clinicians involved in the study are blinded to group allocation (Cook & Sheets, 2011). Not only does this uphold the principle of equipoise, but it also minimises selection bias throughout the trial and other important study processes such as allocation, data collection, intervention delivery and data analysis. Blinding in RCTs is "the practice of preventing study participants, health care professionals, and those collecting and analysing data from knowing who is in the experimental group and who is in the control group, in order to avoid them being influenced by such knowledge" (p. 842) (Akobeng, 2005).

Blinding starts at randomisation, which is the process used to assign patients to a treatment arm in an RCT (Akobeng, 2005). It "allows the principle of statistical theory to stand and as such allows a thorough analysis of the trial data without bias" (p 136) i.e. patients should not be cherry picked but assigned at ***random*** to groups in an unpredictable manner (Hoare, 2010). There are various randomisation techniques available, ranging from simplistic to more complex methods such as

restricted and stratified methods. Simple randomisation (e.g. tossing of a coin) satisfies the random element but can lead to unequal sized groups, whereas stratified randomisation allows the research team to have some control over the size of the treatment groups and also account for certain characteristics that could impact upon the main outcome of the study (e.g. gender or age) (Hoare, 2010). These are known as confounding factors and are introduced into the algorithm as 'stratification variables' (Akobeng, 2005; Brocklehurst & Hoare, 2017; Hoare, 2010; Russell et al., 2011). The net result of these different approaches is that the characteristics of participants in intervention and control arms are as similar as possible and any differences between the groups will have occurred by chance (Akobeng, 2005). Allocation concealment is another process that attempts to minimise bias in RCTs. Here, the researcher is blinded to how participants are allocated to different trial arms during randomisation.

The processes of patient selection and randomisation should be detailed in the study protocol along with the level of blinding (open-label, single blind, double blind or triple blind) and any other processes to be followed such as data collection, data handling and analysis, treatment procedures and study intervention deliveries, handling adverse events, non-compliers and withdrawals (Smith & Noble, 2014). A well-defined protocol, which is registered before the start of the trial, plays an important role in reducing bias and ensures that the research team is clear about the processes to be undertaken (Smith & Noble, 2014).

## 1.3 Measurement in RCTs

Another key element that influences the rigor of an RCT is measurement error and as highlighted above, this forms the basis of this current thesis. Measurement error describes a systematic distortion of the findings of a trial as a result of the type of measures that are used and how the outcome data is collected. As the main aim of clinical research is to improve the health and wellbeing of the population, one of the most important considerations in RCT design is how to capture these changes in health and wellbeing states. The most common method is to take measurements of the health of participants in the trial at baseline and again, at predetermined time points after the intervention has been delivered (Streiner, 2008). The inherent assumption in this model underpinned by the empirical paradigm, is that the outcome measure is only influenced by the intervention i.e. that the measure itself *does not change*, but captures meaningful changes in health states, which are valid, consistent, reliable and free from bias.

### 1.3.1 What is an outcome measurement instrument?

An outcome measure is a tool used to evaluate changes in the participant's health state (Streiner, 2008). They commonly take the form of 'instruments' or 'scales' generally developed by professionals and clinicians to evaluate important traits, behaviours, symptoms and experiences of a treatment, illness, disease or health issue (Streiner, 2008). There are several forms of measurement instruments based on the concept of 'levels of measurement' (Stevens, 1951 as cited in Steriner 2008). These levels relate to the format of the instrument and fall into either binary, nominal, ordinal, ratio or continuous categories. This is an important consideration in trials as the variable format impacts on the sample size of the study (Andrade, 2015).

Interventions are sometimes assessed for efficacy by direct empirical measures based on observation, such as changes in size (e.g. tumour size, weight-loss), survival, or changes in biomarkers or laboratory tests (Frost et al., 2007). However, as healthcare has become increasingly patient-centred it is increasingly important to measure patients' perspectives. This has resulted in the development and increase in the use of patient reported outcome measures (PROMs) (Frost et al., 2007; Marshall et al., 2006; Rothman et al., 2007). The advantage of PROMs is that they incorporate the patient's view about their health state and are often used alongside measurements of clinical change (Frost et al., 2007; Rothman et al., 2007; Turner et al., 2007). This means that they can gather data which cannot always be obtained from direct observation of the clinical condition under scrutiny (Rothman et al., 2007). They involve patients answering questions subjectively about their health status such as recording any symptoms, changes to function, experiences with procedures, satisfaction with treatment and health and care utilisation. They can also capture patient's perceptions of their condition and any impact on their well-being and quality of life (Frost et al., 2007; Rothman et al., 2007; Turner et al., 2007). The disadvantage of PROMs is that they can be more prone to bias, when compared to direct observational measures, as they are subjective and so can be open to manipulation (knowingly or subconsciously) by both the researcher and the participant due to their self-reporting nature.

PROMs can measure simple or complicated concepts (e.g. Health Related quality of life (HRQoL)) (Rothman et al., 2007). Depending on the nature of the measurement in question, they may encompass just one item or be composed of multiple items (Rothman et al., 2007). Generally, PROMs with several items are considered to measure health states, traits or behaviours more precisely than those with just a single item as they are considered to capture several attributes simultaneously rather than just one (Frost et al., 2007).

PROMs tend to be presented as questionnaires that are self-completed by participants, but they may also be completed by a researcher at an interview and assessment visit, especially in less cognitively able populations (Marshall et al., 2006). They can either be a generic measurement of a trait or they can be 'disease specific' and there are advantages and disadvantages to both approaches. With generic instruments, some of the items may not always be applicable to some of the participants that are being exposed to the measure. This contrasts with disease specific approaches, where all items *should* be relevant for the participants. As a result, disease specific instruments tend to be shorter than generic scales, but their disadvantage is that they are not as generalizable as generic scales, which allow for comparisons across studies that have used the same scale (Streiner, 2008).

Several studies use proxy rating outcome measures in addition to PROMs. A proxy rating is an outcome measure that is completed by the participant's representative, relative, friend or caregiver (Magaziner et al., 1997). These are often used in Dementia trials, which will be discussed further below. They are commonly utilised when the information required cannot be directly collected from the participant or when there is a question about the reliability of the participant's response, for example, due to cognitive decline (Magaziner et al., 1997).

RCTs are powered to detect a difference on *one* Primary Outcome Measure (POM), which in a hypothetic-deductive paradigm, relate back to the specific research question being asked by the research team (Andrade, 2015). All other measurements are considered to be secondary. As a result, it is extremely important that researchers conducting trials identify and specify the POM during the design stage of the study and state this clearly in the study protocol (Andrade, 2015; Bialocerkowski & Bragge, 2008). The choice of the POM is based on considerations of validity (see above), clinical knowledge and is commonly informed by earlier studies conducted in the literature (The BMJ, n.d.). The choice is also affected by logistical aspects such as administration, responsiveness, interpretability and readability (Lohr, 2002; Terwee et al., 2007). If the POM is not decided *a priori* then there is potential to increase the risk of a Type I error, as a result of multiple outcomes being analysed simultaneously. It can also increase the risk of a Type II error as the POM should be used to deduce the estimation of the sample size needed for the study to have sufficient statistical power (Andrade, 2015).

### 1.3.2 What is measurement error

Exploring measurement error in Dementia trials forms the basis of this thesis. The precision and accuracy of POMs is crucial to RCTs (given the inherent assumption of stability referred to above) and is particularly important in Dementia trials, as they commonly use more subjective measures of

experience due to the nature of the disease process. However, as highlighted in the preceding section, there is a lot of potential to introduce measurement error when collecting and evaluating outcomes that are subjective in nature. Measurement error can arise in many ways in RCTs and what researchers mean by measurement error varies in the literature and is not always clear. This thesis regards measurement error as a separate form of internal validity, as shown in Figure 2. Both random error and systematic error contribute to measurement error and the internal validity of a study. Within the systematic error element, the biases that are detailed in Section 1.2 can contribute to overall 'measurement error' of the treatment effect and hence, have an impact upon internal validity. This section of the Chapter will now focus on 'measurement error', which refers to a systematic error or bias in relation to the measuring tool or instrument used in the study, as shown in the red box in Figure 2.

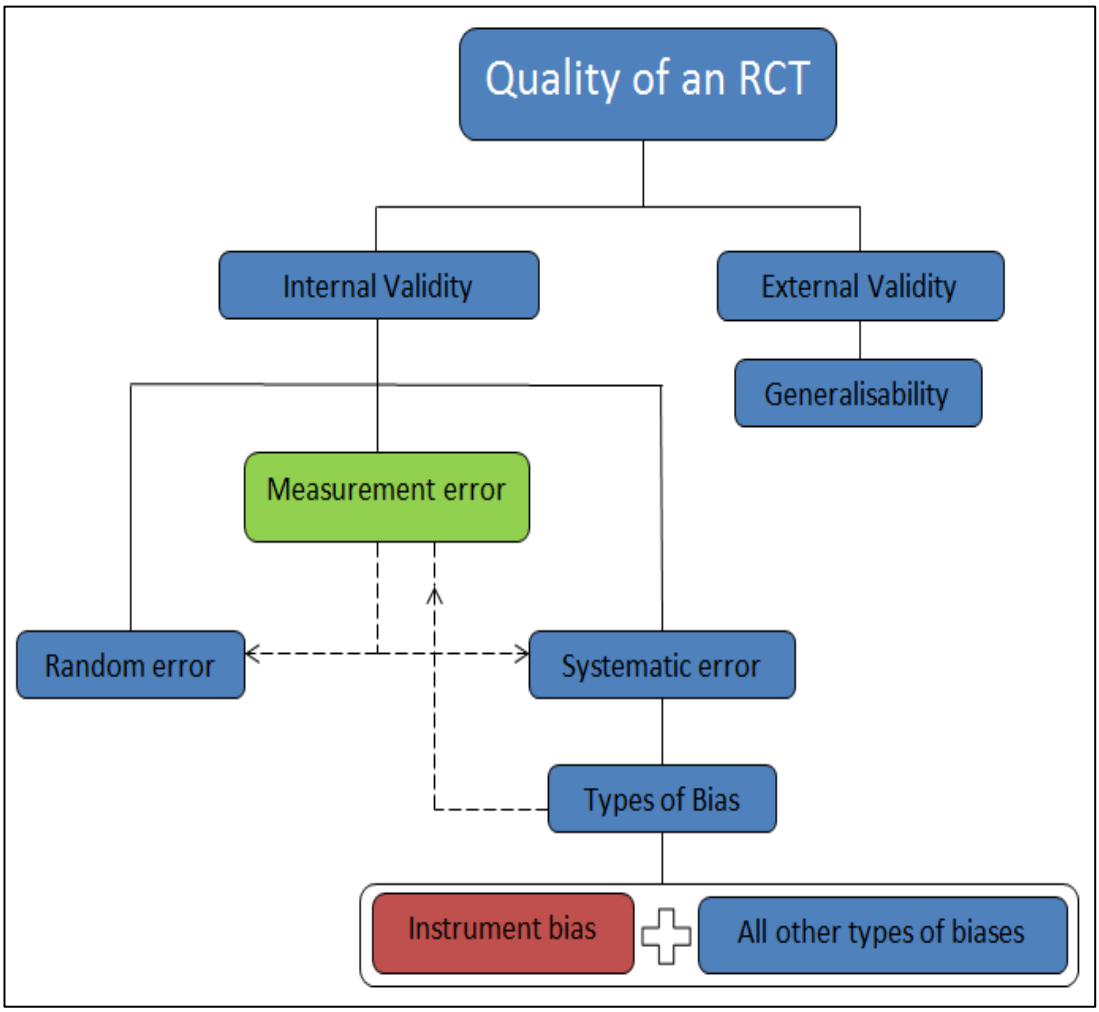


***Figure 2:*** *Components to Evaluate the Quality of a Randomised Controlled Trial*

When undertaking a measurement, there will always be some form of inaccuracy, even with validated tools and instruments (Bartlett & Frost, 2008). This is especially true when several parallel measurements are undertaken on the same individual or object, due to natural variation in the participant, variation in the measurement process, or a combination of the two (Terwee et al., 2007). Measurements that are used daily, for example, scales, temperature gauges, clocks (analogue or digital), speedometers and many more, are generally accepted as having some form of inconsistencies and fluctuations resulting in measurement error (Streiner, 2008). To understand measurement error further, let's take an example of luggage weight when travelling on an aeroplane. Each passenger has an allowance of 'X' Kg. Typically, one will weigh their suitcase before they leave for the airport and then the ground staff will re-weigh the luggage at the airport at check-in using their scales. The measurement readings will likely be slightly different, although nothing has been added or removed in the suitcase. This would be regarded as a difference caused by using different instruments. If you had used the same scales, you would expect the differences in the two readings to be similar but may still expect some variation due to your own imprecise readings.

This raises the question about how much error in a measurement is tolerable. You would hope that the airline's scales are accurate as you don't want them to overestimate the weight, as this may attract excess charges. Equally, you don't want them to underestimate the weight as each plane has a weight limit for safe and efficient flying. As a collective of individual measurements, the error and magnitude might be more of a concern. If a few people are only over their allowance by a small amount, it is not likely to make a difference to plane safety. However, if every passenger goes over their luggage allowance by a large amount then this could have a more significant impact. Thankfully, there is a large window for measurement error in the airline business and it would require a substantive amount to impact upon aviation safety. However, one might argue that in clinical practice individual measurement error is more of a concern and the collective impacts reported in a study may influence the final conclusions drawn about the intervention under scrutiny (Bartlett & Frost, 2008).

To quantify outcome measurement error, we assume that each measurement taken will have a 'true' value, which is the value that would be obtained if a perfect reading was obtained. Each measurement taken will consist of this 'true value' and some amount of error, which can be expressed as: **Obtained measured value = true value + measurement error** (Rothstein, 1985 as cited in Bialocerkowski & Bragge, 2008). Rearranging this gives **Measurement error = True value – obtained measured value**. Unfortunately, the only known variable in the equation is the 'measurement obtained value' and the error value can only be estimated and not calculated directly (Bialocerkowski

& Bragge, 2008). This is done using reliability and validity statistics, which are discussed below (Terwee et al., 2007).

### 1.3.3 Measuring outcome measurement error

In RCTs that use these types of tools or instruments, it is important to understand the extent of this measurement error in order to evaluate the magnitude of this error on the results of the study in question (DeVon et al., 2007; Shrout & Fleiss, 1979). This can be done by obtaining the tool or instrument's psychometric properties. Psychometric properties provide a means of understanding the relative performance of the instrument under question (Bialocerkowski & Bragge, 2008; Turner et al., 2007). All outcome measure instruments should demonstrate sufficiently high psychometric properties so that they can be used with confidence in RCTs, as this forms the basis of the empirical approach (Rosenkoetter & Tate, 2018). Given the ability of RCTs to determine causality, a robust and stable instrument is important in order to attribute causal change. Monkkink et al. (2010) produced the COSMIN checklist (COnsensus-based Standards for the selection of health status Measurement Instruments) that includes a list of the properties that should be assessed in order to provide confidence in a given health related PROM. There are many terms used in the literature: 'agreement', 'reliability', 'reproducibility', 'repeatability' and 'validity' (Bartlett & Frost, 2008). Here, we discuss the two main parameters - reliability and validity.

#### Reliability

The concept of reliability describes a tool or instrument's ability to accurately replicate a measurement over a given timeframe (Frost et al., 2007; Streiner, 2008). This relates to the consistency and repeatability of the measurement tool and as stated above, this is fundamental to the empirical paradigm (DeVon et al., 2007; Rosenkoetter & Tate, 2018). Reliability is often referred to by other researchers and authors in the literature as "*objectivity, reproducibility, stability, agreement, association, sensitivity and precision"* (Streiner, 2008), but it is only one element of measurement error (DeVon et al., 2007). There are three main types of reliability: test-retest reliability, inter-rater and intra-rater reliability (highlighted below) (Bialocerkowski & Bragge, 2008; Gwet, 2014; Rosenkoetter & Tate, 2018). All are evaluated using the intraclass correlation coefficient (ICC) (Fisher, 1992 as cited in Streiner 2008; Shrout & Fleiss, 1979;). The calculation of the ICC is given by the participant variability divided by the total variability using variance calculations (Streiner, 2008):

$$\textbf{ICC (Reliability coefficient)} = \frac{\textbf{participant variability}}{\textbf{total variablility}}$$

The value of an ICC will fall somewhere between 0 and 1, the higher the value, the better the reliability of the test and the lower the value of the co-efficient, the less reliable the test i.e. a 0 would indicate 'no reliability' and a 1 would indicate 'perfect reliability' (Bartlett & Frost, 2008; Streiner, 2008). However, interpretation of the magnitude of the coefficient is not straightforward. (Frost et al., 2007; Streiner, 2008) suggest that an ICC lower than 0.70 is not a sufficiently robust enough for use in RCTs, but as highlighted by Streiner (2008), the performance of the measure is also reliant on the sample size of the study, in which the measure will be used. There is no general agreement for the calculation of a sample size that is required to obtain a stable ICC: there are suggestions in the literature that a minimum sample of either 200 or 400 participants is required (Charter, 1999; Frost et al., 2007) and broadly, the lower the reliability of the measure, the larger the sample size that is required (Kraemer, 1979 as cited in Steriner 2008).

Test-retest reliability commonly refers to tests where there are no observers involved, such as self-completed questionnaires, an online survey or automated telephone questionnaire completed by the participant alone i.e. there is no 'observer' effect created by the researcher (Portney & Watkins, 2000 as cited in Bialocerkowski & Bragge, 2008). However, variability can be introduced by the participant themselves, over and above the variation caused by underlying changes in the disease state being measured. Here, the same measurement instrument is administered under the same conditions to the same participant on two or more different occasions and the reliability of the test will indicate the consistency of the measurement tool, when the traits, symptoms or behaviours of the test being evaluated have not changed (Bialocerkowski & Bragge, 2008; Streiner, 2008).

Inter-rater and intra-rater reliability coefficients are also referred to as inter-observer and intra-observer reliability. These terms relate to studies which use different raters (e.g. clinicians or researchers) obtaining measurements on a specific tool or instrument (Portney & Watkins, 2000 as cited in Bialocerkowski & Bragge, 2008). This type of error forms the basis of this thesis, in relation to Dementia trials. The difference between these two terms relates to whether the same or a different rater is being used to measure the effect (Gwet, 2014). Int**er-**rater reliability is evaluated when two or more different raters obtain measurements on the same patients (Portney & Watkins, 2000 as cited in Bialocerkowski & Bragge, 2008). The coefficient obtained estimates the amount of error when different raters take a measurement by obtaining an estimate of the probability that the different raters will obtain the same results on the same patients (Bialocerkowski & Bragge, 2008). Ideally there should be little variation amongst raters' scores on the same patients, although there will always be some natural variation across raters' scores (Bialocerkowski & Bragge, 2008). A small value for the ICC

demonstrates good reliability, whilst a large coefficient suggests that the raters are not administering, scoring or obtaining the measurement questionnaire consistently, when compared to one another (Bialocerkowski & Bragge, 2008).

Another form of the ICC coefficient is the int**ra**-rater reliability (Streiner, 2008). Rather than identifying variations between observers, the intra-rater coefficient assesses the consistency of measures taken over time by the same rater (Portney & Watkins, 2000 as cited in Bialocerkowski & Bragge, 2008; Streiner, 2008). It involves one rater evaluating measurements on two or more occasions on the same participant (Bialocerkowski & Bragge, 2008). Again, there will be natural variation from the observer and expected variation in the participant, but the ICC should still be high enough to indicate that there is not too much error present. The issue with intra-rater is the time period in-between the different measurement points; too long and several factors in the participant and observer may have changed. If too short, the researcher or participant may carry memory of performance from the previous observation into the current measure assessment known as repeat testing bias (Bialocerkowski & Bragge, 2008).

### Validity

Previously in this Chapter we discussed validity in terms of internal and external validity (see Figure 2 in section 1.3.2). Along with reliability, the validity of an *outcome measure instrument* is a psychometric property pertaining to whether the instrument is measuring what it aims to measure and not another concept. As a result, this form of validity is often termed 'construct' validity (Streiner, 2008).

For some 'direct' clinical observation assessments or measures such as survival, the validity of the test isn't likely to be called into question (Streiner, 2008). However, when the underlying phenomena under investigation is more complex, such as HRQoL, anxiety and depression, it is more difficult to assess what *exactly* is being measured by the instrument. As a result, a process is commonly undertaken to demonstrate the validity of the instrument (Streiner, 2008). This is usually done by comparing the measure to another instrument (or several others) that is evaluating a similar concept; in this process, the correlation between the different measures gives an indication of whether the new instrument has sufficient *construct validity* (Streiner, 2008).

### Relationship of Reliability and Validity in RCTs

Reliability and validity are two distinct psychometric properties of a PROM and measurement error can affect both (Frost et al., 2007; Streiner, 2008). If an instrument demonstrates good reliability,

this does not mean that the instrument is valid (Bialocerkowski & Bragge, 2008; DeVon et al., 2007; Frost et al., 2007). In a similar manner to bias, errors caused by inconsistency in the instrument's reliability and validity are considered to be on a spectrum i.e. neither completely reliable/valid or unreliable/invalid (Frost et al., 2007). The ICC of both properties of the PROM need to be high enough for the outcome measurement tool to be used in RCTs (Bialocerkowski & Bragge, 2008). Otherwise, causation could be misattributed to the intervention, when the phenomena could be as a result of the inconsistent properties of the POM. In most studies, this threshold is set to be a minimum of 0.70 (Post, 2016).

Poor reliability impacts on both the sample size and study power in an RCT. A power calculation (also referred to as a sample size calculation) is undertaken *a priori* to determine the number of participants that are required to detect a clinically meaningful difference, if one is present. The calculation consists of the following components: intended study power (typically set at 0.8 or 0.9 for 80% or 90% power), the significance level (generally 0.05) to avoid a Type I error and the estimated effect size of the intervention. These three elements determine the number of participants required for a trial (sample size). Any one of these variables can be calculated if the others are known and therefore each parameter is interdependent. Increasing the number of participants in the sample increases the power of the study (Leon, 2007; Leon et al., 1995). However, increasing the sample size might not be feasible in some trials as the resources may not be available to recruit extra participants or the required population size in the cohort of interest might not be large enough due to limited prevalence (Leon, 2007). As a result, increasing the ICC can be considered as an alternative and more reliable method to increase the power of the study (Leon et al., 1995). This is because the effect size (Cohen D calculation of effect size) is obtained by a 'signal-to-noise ratio' where the signal is 'the magnitude of the difference between the means of the two groups' and the 'noise is variation among the experimental participants' (Cohen, 1988; Leon et al., 1995). Therefore, reducing this 'noise' (measurement error), decreases the variability around the point estimate (Leon, 2007). In turn, this increases the power of the study and/or decreases the sample size (Leon, 2007). Perkins et al. (2000) statistically modelled the relationship between the reliability of the outcome measure and sample size. Their results showed that if the reliability of the measure can be improved from 0.7 to 0.9 then a 22% decrease in the sample size required can be achieved without reducing study power by (Perkins et al., 2000).

Despite its importance, measurement error tends to receive less attention than other aspects of research design and becomes increasingly critical with measures that are not objective in nature

(Leon, 2007). Details concerning the reliability of measures are infrequently reported with study results and this has particular ramifications for trials that rely on subjective measures, like Dementia trials (Kobak, 2010). For example, a review carried out by (Mulsant et al., 2002) on studies of depressive disorders conducted between 1996 and 2000, found that factors that can influence measurement error are poorly reported. Several aspects of rater characteristics, such as the number of raters used and rater training, were not reported. The review also found that only 22% of single centre and 14% of multi-centre RCTs reported IRR coefficients (Mulsant et al., 2002). Given the potential of measurement error to undermine the empirical basis of the trial paradigm, RCTs should use standardised methods of scoring on measures and researchers should be trained on these processes (Kobak, 2010). Indeed, some argue that it is necessary to evaluate and monitor raters' scores throughout trial processes to minimise variability and bias (Gaur et al., 2009).

### 1.3.4 Sources of instrument specific outcome measurement error

Instrument specific outcome measurement errors are biases introduced as a result of the instrument design and its use. In a similar manner, researcher and observer effects and can contribute to bias as a result of response and recall bias.

#### Response bias

Response bias describes how a participant's perspective and knowledge of a given phenomenon, as well as their motivation for completing the assessment, can have an effect on the responses given. A fear of judgement by the assessors can directly influence the participant's response, artificially increasing or decreasing the score provided by the participant. A participant's desire to complete the assessment as quickly as possible can also impact on the response that is given.

Depending on the sensitivity of the questions being asked, there could be a difference between a participant's true answer and how they wish to be portrayed. This demonstrates the concept of social 'desirability and faking good' and conversely 'deviation and faking bad' (Streiner, 2008). These concepts can either be adopted to avoid judgement or to achieve a particular status (Streiner, 2008). This can lead to a 'hello-goodbye' effect, where participants may portray themselves more negatively before an intervention in order to meet the eligibility criteria of a trial and then portray themselves more positively post-intervention (Streiner, 2008). This type of bias can be introduced by participants inadvertently, to avoid judgment, or deliberately in order to be included in a particular trial. This phenomenon can be reduced by concealing the nature of the trial, but this raises important ethical considerations around full disclosure (Streiner, 2008).

Other measurement errors can be introduced when questionnaires are designed with only positive or negative answers (Streiner, 2008). Use of both positively and negatively coded responses can help identify and minimise this (Streiner, 2008). Respondents can also be end-averse, avoiding the response items at extreme ends of scales (Streiner, 2008). Again, these can be minimised by the researcher paying particular attention to the types of end statements used (Streiner, 2008). Binary yes/no responses can also limit the ability of participants to give accurate honest answers if their answer lies truly on a scale (Streiner, 2008).

### Recall bias

In many studies, participants are often asked to reflect on phenomena retrospectively i.e. **recall** details of the event(s) such as their experiences, symptoms and behaviours; the severity of them; the number of occurrences and the impacts of them and emotions felt. This all involves memory (remembering what happened), making estimations of the event(s) (how many times it happened and over what time period it had an affect) and also making inferences of the event(s) (the severity and impacts) (Streiner, 2008). Completing tools and instruments that have a substantive retrospective element can be cognitively draining and difficult for some participants, introducing bias at several stages of the research process (Streiner, 2008).

This can particularly be an issue in Dementia trials. Participants may amplify or down-play certain events and their perspectives can be influenced by their individual experiences and their ability to accurately recall these experiences (Pannucci & Wilkins, 2010). It is easier for participants from the general population to remember events that have occurred more recently rather than further in the past (Indrayan, 2012). The implications of this are that the events may appear exaggerated or go under-reported depending on when they are asked, which could impact upon the study findings. Researchers should consider this when deciding on time points for primary outcome collection and other data collection (Pannucci & Wilkins, 2010).

### Instrument design

Many of the biases described above can be mitigated by the design of the instrument; this is why there should be several design and development steps incorporated in producing measurement instruments. This is particularly important in population cohorts that are more vulnerable to these types of bias. The process should involve theoretical explication, qualitative exploration, quantitative confirmation, and psychometric support (Turner et al., 2007). Consideration should be given to vocabulary, comprehension level, balance of positive and negative loading and questionnaire length. In relation to length, consideration should be given to how to measure the intended concept and the

questionnaire should contain enough items to cover all of the content areas required, without being too long. How the question is asked and how the participants are given the opportunity to answer the question are also key. Response options can cover a full range of responses but still not contain an individual's exact answer, thus requiring them to choose the closest appropriate answer, which in turn can introduce some loss of precision (Streiner, 2008). Precision can also be lost with the participant's interpretations of the response options. For example, 'often', 'usually', 'sometimes' and 'frequently' can vary in interpretation and ranking from person to person.

### Environmental factors

There are several ways of administering outcome measures in different environments with advantages and disadvantages of each. Face-to-face is the most commonly used method of obtaining health measurements but others include telephone, mailing, computer assisted, random digit dialling, e-mail questionnaires and the use of websites (Streiner, 2008). There is an extensive amount of literature and research covering the relative pros and cons of the different methods of administering questionnaires and surveys which summarises the impacts of each method on the measurement and potential measurement error which is beyond the scope of this thesis.

Participants' responses can be affected by environmental distractions, Bialocerkowski & Bragge (2008) suggest that errors can be reduced during assessments by giving the participant a quiet environment with limited distractions to facilitate completion of the questionnaire. It is also said that members of a participant's family can influence their responses so, where possible, participants should complete assessments or questionnaires without the presence of family, friends or care givers (Bialocerkowski & Bragge, 2008).

### Researcher bias

Whilst a lot of research focus is around scoring differences and ICCs, less attention has been paid to the quality of the interview carried out by researchers to collect outcome data (Kobak, 2010). Researchers may be successfully trained on scoring outcome measures and hence achieve good reliability coefficients, but a factor that can be overlooked is that the instrument items and overall scores are generated from the responses given to the researcher from a participant during interviews (Pannucci & Wilkins, 2010, Streiner, 2008). The way in which researchers obtain this data through questioning may contribute to rater variation across measurements in a study (Davis et al., 2010; West & Blom, 2016). This is particularly the case in Dementia studies, some authors have suggested that several Alzheimer's RCTs that had a robust research design may have 'failed' due to poor inter-rater reliability, poor interview quality and rater bias (Kobak, 2010). There are a variety of suggestions of

ways to minimise this type of bias in the literature: for example, ensuring all researchers collecting data are blind to participant allocations (Allen, 2002; Pannucci & Wilkins, 2010); formatting interview questions to ensure influence isn't introduced and to ensure that researchers are trained properly in data collection processes (Malone et al., 2014; Streiner, 2008).

Although research has been conducted on interviewer or researcher effects on participant responses (Davis et al., 2010; West & Blom, 2016), there is little literature which addresses any effect of the consistency of the researcher collecting outcome data between visits and the impact this can have on the measurement and participants' responses. Kobak (2010) suggests that researcher consistency may introduce potential biases on outcomes. He hypothesised that those who had the same researchers may have ratings that are biased as the rater has prior knowledge of the participant's health state at baseline or the previous assessment and therefore may score the measure based on this (Kobak, 2010). However, this is insufficiently evaluated.

The current thesis will explore whether there is any difference of the same researcher attending follow-up interviews and collecting outcome measures as opposed to different ones which will give an indication of whether a possible researcher effect on the outcome measure is present. Data from a previously conducted RCT with multiple time points evaluating QoL outcomes in a Dementia population will be utilised.

People with Dementia generally have a lot of contact with health care professionals (e.g. nurses, carers etc.) and may not necessarily have consistency with these persons ("Health and Social Care professionals", n.d.; Prince, 2016). Introducing a researcher into their environment on top of the numerous health care workers may add more stress for them. Continuity of researchers may be therefore important for this population as people with this disease may value stability (The Good care group, n.d.). Additionally, certain types of measurement error may be more likely in these studies as some of the characteristics and symptoms of the disease, which are discussed in Chapter Two in more detail, are likely to contribute to certain types of measurement error, for example memory loss may introduce recall bias.

This Chapter has noted the importance of RCTs to EBM and their reputation as the 'gold standard' of research. The issue of quality of RCTs was raised and the many biases that are prevalent across RCTs was highlighted. The key importance of the POM was considered and in particular PROMs were discussed. Measurement error of these PROMs and methods of measuring measurement error

was explored. Instrument specific biases were mentioned with and attention on researcher bias. The question around the impacts of researcher consistency on measurements were raised. The next Chapter will cover a brief introduction to Dementia and the impacts on the patient's (and their carers) quality of life. It will explore the issues of measuring quality of life in this population, pertinent to the points raised above.

# Chapter Two – Introduction to Outcome measures to assess QoL in Dementia

Dementia has become an increasingly common medical condition over the last few decades and is now a major focus in health and social care. The prevalence of individuals in the United Kingdom (UK) with Dementia in 2019 is estimated to be around 885,000 and predicted to rise to 1 million by 2024 and to 1.6 million by 2040 (Wittenberg et al., 2019). In the UK, currently one sixth of people over the age of 80 have Dementia and approximately 40,000 people under the age of 65 have the disease ("Facts for the media", n.d.). Dementia patients or those with cognitive problems account for 70% of the population in care homes across the UK ("Facts for the media", n.d.). Worldwide, there are approximately 47 million people living with Dementia, which is predicted to grow to 132 million by 2050 (World Health Organisation, 2017).

Dementia is a broad umbrella term for a group of diseases characterised as “different brain disorders that trigger a loss of brain function” ("Facts for the media", n.d.; World Health Organisation, 2017). The medical dictionary definition of the condition states that an illness will be classed as Dementia “when both memory and another cognitive function are each affected severely enough to interfere with a person’s ability to carry out routine daily activities” (Merriam-Webster, n.d.). The most common form of Dementia is Alzheimer’s disease (AD) with around 60 - 70% of Dementia cases being classed as AD (World Health Organisation, 2017). Other common types of Dementia include Vascular Dementia, Dementia with Lewy Bodies and Frontotemporal Dementia (World Health Organisation, 2017). There are also many other rarer types such as Young-onset Dementia, Alcohol-related brain damage and HIV-related cognitive impairment ("Types of Dementia", n.d.; World Health Organisation, 2017). In around 10% of cases patients have a combination of Dementia forms recognized as ‘Mixed Dementia’ ("Facts for the media", n.d.; World Health Organisation, 2017). The conditions are progressive and are generally classed into either early stage, middle stage or late stage Dementia (Wittenberg et al., 2019; World Health Organisation, 2017).

Although the exact cause of Dementia is unknown, there are certain factors which can increase a person’s risk of developing the disease. Certain demographics (e.g. age, gender, and ethnicity), lifestyle choices (e.g. unhealthy diet, lack of physical activity, excessive alcohol abuse and smoking) and other medical conditions (e.g. cardiovascular disease, high blood pressure, diabetes, depression and heart disease) are all recognized as Dementia risk factors (“Risk factors for Dementia”, n.d.; World Health Organisation, 2017; Wu et al., 2016).

Whilst the severity of the condition progresses at different rates, the disease is a progressive disease that eventually leads to death and each year the disease contributes to a large proportion of deaths in the UK (Office for National Statistics, 2019; World Health Organisation, 2017;). In 2018, the disease was the leading cause of death in the UK (Office for National Statistics, 2019). There are a number of management strategies to reduce the symptoms associated with the disease (i.e. cognitive behaviour, insomnia, aggression, depression) ("Treatments", n.d.). These include pharmaceutical medications, person-centred care, talking therapies and other alternative therapies such as aromatherapy, massage, bright light therapy, Cannabidiol (CBD) oil, coconut oil and transcutaneous electrical nerve stimulation ("Treatments", n.d.). However, there are currently no treatments or medications that can cure the disease ("Facts for the media", n.d.; Brod et al., 1999; Prince et al., 2014). The cost of Dementia to the National Health Service (NHS) in the UK is stated to be approximately £26 Billion per annum and is predicted to continue to rise further (Prince et al., 2014). As a result, there is high demand for research into the disease (Jopling, 2017).

While the different Dementia disorders as a group share similar characteristics of irreversible cognitive decline and neurological problems, the symptoms of the diseases differ between types of Dementia can vary on a case-by-case basis (Banerjee et al., 2009). In general, patients tend to suffer with symptoms such as memory loss, confusion, reduced understanding, effect on judgement and reasoning, difficulties with speech and language, difficulties learning new tasks and lack of motivation (Association, 2013; Wu et al., 2016). The disease also affects one's ability to conduct activities such as domestic tasks and self-care (Association, 2013; World Health Organisation, 2017; Wu et al., 2016). Behavioural and personality changes such as intensified emotions including aggression, agitation, less patience and anxiety are also common symptoms (Cerejeira et al., 2012). These symptoms associated with the disease can be degrading, troublesome and worrying for the individual diagnosed and affect their daily activities, therefore people with Dementia tend to experience a reduction in quality of life (QoL) (Lewis et al., 2014; Steeman et al., 2007). Quality of life is described as "the standard of health, comfort and happiness experienced by an individual or group" ("Quality Of Life", n.d.).

The symptoms of Dementia result in patients requiring a significant amount of care and support (World Health Organisation, 2017). Most carers of People with Dementia (PwD) are informal and tend to be family members, particularly their spouse or children (Zwaanswijk et al., 2013). Informal carers typically do not have the specific skill set, guidance or training of a professional carer (Zwaanswijk et al., 2013) and the role can be straining both mentally and physically for them (Brod et al., 1999; Zwaanswijk et al., 2013). The behavioural changes in patients with the disease may be

especially distressing for the carer particularly as they tend to be 'out of character' for them (Spector et al., 2016). Compared with carers of physically impaired older people in other disease areas, Dementia caregivers tend to experience the negative side effects to a higher extent, reporting higher stress levels, more mental health issues, less time for other family members and other social relationships and more work-related problems (Bremer et al., 2015; Dow et al., 2018; Schulz & Martire, 2004; Thomas et al., 2015). Research into the impact on informal carers has shown that the caregiving tasks associated with caring for PwD results in substantial burden and the responsibilities of caring for a PwD can have a negative impact on their QoL (Dow et al., 2018; Merlo et al., 2018; Zwaanswijk et al., 2013).

The patient-carer relationship can also be complicated. As many PwD-carer relationships are formed before a diagnosis is made, both the PwD and carer not only have to come to terms with the diagnosis, but are both required to adapt to their new roles within this relationship (Spector et al., 2016; Quinn et al., 2009). It appears that the quality of the patient-carer relationship impacts upon the QoL of both the PwD and their carer, with those reporting a better relationship, scoring higher on respective QoL measures (Marques et al., 2019, Quinn et al., 2009; Woods et al., 2014). Equally, where more care is required, this is likely to lead to a reduction in the carer's QoL (Quinn et al., 2009). As a result, the importance of considering QoL in Dementia research should not be underestimated and QoL is an important area of study (Ready & Ott, 2003). Interventions to improve QoL in Dementia patients and carers have become increasingly examined over the last few decades (Lyketsos et al., 2003; Missotten et al., 2008; Mougias et al., 2011). As a result, to assess the impacts of these interventions, the measurement of QoL in Dementia has become a key focus (Naglie, 2007; Woods et al., 2006).

One of the main issues with measuring QoL is that it is not always clear what is meant by the term (Bosboom et al., 2012; Missotten et al., 2008). Definitions vary and the constructs encompassing these definitions differ. The World Health Organization (WHO) broadly defines QoL 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" (World Health Organisation [WHO], n.d.). In this context, QoL is more than just the health status of an individual and is not classified by a person either having or not having a disease; nor is it measured by the severity of symptoms or the presence of symptoms (Millenaar et al., 2017; World Health Organisation, [WHO] n.d.). The concept relates more to the individual's view of their own health state, happiness and sense of contentment (Medvedev & Landhuis, 2018).

A largely influential framework for designing QoL measures in Dementia is the work of Lawton and many Dementia specific QoL instruments have subsequently been based on his research (Ettema, Dröes, et al., 2005; Naglie, 2007; Thompson & Kingston, 2004). QoL consists of both objective factors and subjective factors and the evaluation of both is required to appropriately measure it (Lawton, 1983 as cited in Thompson & Kingston, 2004). It is acknowledged as a complex phenomenon with various factors that consist of both physical and psychological domains along with social domains such as PwDs' relationships, environment, support, values and coping styles (Banerjee, 2006; Bosboom et al., 2012; Beerens et al., 2013; Mougias et al., 2011; Walker & Lowenstein, 2009). Many of these factors contribute to the measurement of QoL, but what affects one person's QoL might not affect another's. In addition, the relative importance of each factor to an individual may vary (Thompson & Kingston, 2004). It is therefore important that QoL measurement tools in Dementia cover a range of these constructs and that the opinions and perspectives of PwD are incorporated (Bowling et al., 2015). Instruments that measure QoL need to consider the perspective of the person evaluating the QoL, as scores may vary depending on whether they are evaluated by a health professional, relative or friend of the PwD (Bowling et al., 2015).

For PwD, measuring QoL has added complexity because the characteristics and symptoms of the disease can vary and contribute to measurement error during assessments (American Psychological Association, 2012; Kobak, 2010). This can add to measurement error (see Chapter One). In addition, the measurement of QoL can be cognitively demanding, requiring an assessment of one's situation using both short-term and long-term memory (Black et al., 2012). The neurological impairments of the disease can make this particularly difficult for PwD, potentially leading to measurement error (Addington-Hall & Kalra, 2001; Banerjee et al., 2009). Many argue that the loss of insight and awareness that PwD experience, impacts on their ability to evaluate their own QoL (Ready & Ott, 2003). However, some question how relevant the PwD's lack of insight is (Trigg et al., 2011). As QoL is largely a subjective concept, they would argue that this is still a valid question. If PwD are unaware of their Dementia-related symptoms and rate their QoL higher than other's might, they are still providing an account that reflects their own perception of the disease and its impact on their daily lives (Brod et al., 1999; Trigg et al., 2011). However, other symptoms such as fatigue, loss of concentration, lack of understanding and other problems add weight to the idea that their QoL evaluation is difficult to communicate (Addington-Hall & Kalra, 2001; Banerjee et al., 2009; Kobak, 2010). This has led some researchers to question whether a PwD can self-report without introducing considerable measurement error (Ready & Ott, 2003; Trigg et al., 2011).

The concerns around the reliability of a PwD's self-reported QoL, raises questions about who should undertake this; whether the instrument should collect data from the participants themselves or use a 'proxy' measure from an assigned person (usually a family member or caregiver) (Naglie, 2007; Ready & Ott, 2003). Many argue that proxy measures should be used, but there remains uncertainty around the reliability and validity of proxy ratings for QoL assessments of PwD (Bowling et al., 2015). However, previous studies have found a lack of agreement between the two with many suggesting that proxy ratings tend to be lower than self-reported ratings (Bowling et al., 2015; Huang et al., 2009; Naglie, 2007; Naglie et al., 2006; Ready et al., 2004; Sands et al., 2004; Trigg et al., 2011). While this could be due to the unreliability of the PwD ratings many suggest otherwise and argue that proxy ratings should not be used in substitute of self-ratings (Bowling et al., 2015; Römhild et al., 2018).

There are many QoL outcome measures that have been developed for PwD (Naglie, 2007; Ready & Ott, 2003; Trigg et al., 2011; Yang et al., 2018). The Quality of Life in Alzheimer's Disease (QOL-AD) (Logsdon et al., 1999), Dementia Quality of life Questionnaire (DQOL) (Brod et al., 1999), Alzheimer's Disease Related Quality of Life (ADRQL) (Rabins et al., 1999) and Dementia Quality of Life questionnaire (DEMQOL) (Smith et al., 2005) are frequently cited and used measures in the literature (Bowling et al., 2015;Yang et al., 2018). Other examples include the QUALIDEM (Ettema et al., 2007), the Quality of life in late-stage Dementia (QUALID) (Weiner et al., 1999), observational Dementia care Mapping (DCM) (Chenoweth & Jeon, 2007), the Bath Assessment of Subjective Quality of Life in Dementia (BASQID) (Trigg et al., 2007), Cornell-Brown Scale for Quality of life in Dementia (Alexopoulos et al., 1988), the Quality of Life Assessment Schedule (QOLAS) (Selai et al., 2001), Pleasant Events Schedule in Alzheimer's Disease (PES-AD) (Logsdon & Teri, 1997), and others (Bowling et al., 2015). With a wide range of measures available it can be difficult for researchers to know which measures are appropriate to use. This is important in the context of trial design (Leon, 2007).

As described in Chapter One, a variety of psychometric properties are used to evaluate the reliability and validity of outcome measures and a number of studies have been conducted on the QoL measures used in Dementia research. The validity of these scales has been assessed in several ways (Ready & Ott, 2003). Some have been validated by comparing them against other outcome measures of disease severity, depression, mood, activities of daily living or against generic QoL instruments (Ready & Ott, 2003). The QUALID scale demonstrated good validity in one study by comparing scores with a depression outcome measure (Ready & Ott, 2003). Another study demonstrated convergent validity on the D-QoL by using subscale scores and comparing these two scores on the Geriatric

Depression Scale (Brod et al., 1999). Another study used visual analogue scale scores to validate the Cornell-Brown scale (Ready & Ott, 2003).

As described in Chapter One, the reliability of measures can be assessed using a number of different properties: internal consistency, test-retest scores, sensitivity to change, inter-rater reliability and intra-rater reliability. Previous studies have demonstrated good to excellent internal consistency scores for Dementia specific QoL measures. For example, the ADRQL, the Cornell-Brown scale, the QUALID and the QOLAS measures had ICC Cronbach alpha coefficients of 0.80, 0.81, 0.77 and 0.78 respectively (Ettema, et al., 2005; Ready & Ott, 2003). Test-retest reliability has also been shown to be moderate to high in some of these measures. For example, re-test correlation coefficients on the D-QOL range from r = 0.64 to r = 0.90 and on the QUALID a coefficient of r = 0.81 was obtained (Ettema, et al., 2005; Ready & Ott, 2003). Other studies demonstrated the subscales of the DCM to have poor to moderate test-retest reliability with ICCs ranging from r = 0.33 to r = 0.55 (Ettema, et al., 2005; Fossey et al., 2002). These findings potentially challenge some of the assertions, highlighted above, about the difficulty of evaluating QoL in PwD. These measures appeared to demonstrate good internal consistency and suggest that participants are responding consistently to the different elements inherent in the QoL measures.

The QoL-AD is one of the most used measures to evaluate QoL in Dementia RCTs and many researchers have recommended its use (Harrison et al., 2016; JPND research, 2015). As a result, it is one of the most utilised and researched QoL measures in Dementia (Bowling et al., 2015). Several studies have shown high internal consistency for both participant and caregiver reports, with ICC coefficients ranging from 0.80 to 0.90 (Logsdon et al., 1999; Logsdon et al., 2002; Selai & Trimble, 1999). Other studies have reported the measure's test-retest coefficients to have 'good reliability': an ICC of 0.79 for participants and 0.92 for carers was obtained in one study (Logsdon et al., 1999). As the measure is commonly completed by both PwD and their caregiver, agreement between the participant-carer scores for the scale has also been evaluated. For example, a correlation coefficient of r = 0.40 was obtained, an ICC for absolute agreement of 0.19 and an ICC for consistency of 0.28 (Logsdon et al., 1999; Logsdon et al., 2002).

As discussed in Chapter One, a large source of variability contributing to measurement error is the way the that outcome data is collected (Kobak, 2010). Commonly, Dementia measures are collected via interviews with the participant and the consistency and variation of Dementia specific QoL measures in terms of rater reliability is an important psychometric property to evaluate. Common

psychometric properties here include intra-rater and inter-rater reliability coefficients, both described in Chapter One (Section 1.3.3). Several studies have demonstrated that many Dementia-specific QoL measures have good levels of inter-rater reliability: ICC coefficients of 0.80 for the DCM and 0.90 on the Cornell-Brown scale (Ettema, et al., 2005; Ready & Ott, 2003). The QOL-D's ICC varies from 0.63 to 0.93 and the QUALID has demonstrated an ICC coefficient of 0.83 (Ettema, et al., 2005; Ready & Ott, 2003). However, other studies have found these measures to have low levels of inter-rater reliability. For example, one Dementia RCT measured the reliability of raters on the ADAS-cog and found that reliability was poor with a low ICC coefficient of 0.08 being obtained (Gaur et al., 2009).

Dementia QoL studies typically take place over a long period of time with several follow-up assessments which can affect researcher variation in relation to the quality of the interview and inter- and intra-rater reliability (Connor & Sabbagh, 2008; Kobak, 2010). The lengthy nature of the studies is another contributing factor to the difficulty in measuring PwD responses, as their symptoms generally deteriorate with time. The progressive nature of the disease means participants are likely to further decline across the study period and this may also influence the researcher's expectations of changes to QoL scores over time. This may lead to researchers subconsciously scoring QoL lower at follow-up time points, especially if they have prior knowledge of the participant's previous state (Kobak, 2010).

The long study length also increases the potential for changes within the study team, meaning that different researchers will attend at subsequent follow-up visits. Many Dementia studies have used multiple researchers to carry out visits to collect data at different time points (i.e. REMCARE (Woods et al., 2012), WHELD (Whitaker et al., 2014), iCST (Orgeta et al., 2015) and Challenge DEMCARE (Moniz-Cook et al., 2017)). For example, the Well-Being and Health for People with Dementia (WHELD) study collected data at baseline and undertook a nine-month follow-up across 16 sites in England using multiple researchers (Whitaker et al., 2014). Equally, the individual Cognitive Stimulation Therapy (iCST) trial collected data at baseline, 13 and 26 weeks across four UK sites using several different researchers at each (Orgeta et al., 2015). During a study a participant might have one researcher conducting assessments or may encounter several different researchers.

The impacts of using multiple raters across time points has been little researched and forms the basis of this thesis. This is important as the researcher-participant relationship and the effects of rapport on interviews and assessments is discussed widely in the context of a qualitative paradigm, yet there is often a lack of consideration given to the impact this may have on RCTs, which are underpinned by the empirical notion that changes can be measured objectively (Doody & Noonan,

2013; Guillemin & Heggen, 2009; Pitts & Miller-Day, 2007). Many qualitative researchers would argue that the relationship and rapport between researchers and participants is important for collecting good quality data (Guillemin & Heggen, 2009; Pitts & Miller-Day, 2007). For example, Hill & Hall (1963) concluded that better rapport is associated with higher reliability of the interview data. A recent paper by (Bell et al., 2016) highlights the importance of establishing rapport in quantitative assessments to encourage participants to respond more openly and honestly.

Relationship and rapport are generally built across time through several encounters (Pitts & Miller-Day, 2007). Given the nature of Dementia, these interactions with researchers could be significant. The relationship between the researcher and participant may influence the assessment in several ways. It could be hypothesised that good rapport may promote answers that better reflect the true feelings of the participant, their perceptions and current experience of the disease under question. In contrast, poor rapport may result in the participant providing a response that does not provide a true reflection of their experience or a very limited response. It could be argued that motivation may be lacking in those participants who have a bad relationship with their researcher and less interaction between the two could result in missing data or certain observations not being picked up. However, the impact of good or bad rapport on QoL quantitative assessments is unknown with some research suggesting that too much rapport can influence the interview process in a biased manner (Hill & Hall, 1963; Miller, 1952). Regardless of the question of the optimum level of rapport and the effects that good and bad rapport have on data and outcome measures, researcher continuity is little researched in empirical studies of Dementia. Some researchers have stated that researcher continuity can impact upon retention rates in Dementia trials, suggesting that good continuity "enhances trust and rapport" amongst participants and researchers (Miller, 1952).

Researcher-participant relationships can be complex and the potential to achieve good rapport can be affected by many factors such as interview administration, interview technique or interviewer behaviour (Bell et al., 2016). It has also been said that the researcher and participant's gender can influence rapport and relationship development in the research context (Stahl, 2016) with some research indicating that good rapport can be built more easily between researchers and participants of the same gender (Williams & Heikes, 1993). The term gender here indicates the biological sex of the researcher or participant. Whilst gender was traditionally considered as a binary concept, it is becoming a highly contested area, which goes beyond the scope of this thesis. For the purpose of the analyses that follow, we simply class biological sex into one of two categories; male or female.

Not only does gender play a part in rapport building and relationship development but the gender of the researcher is said to impact upon participant responses (Pollner, 1988; Thurnell-Read, 2016; Williams & Heikes, 1993). It has been suggested that participants are more likely to reveal information to female researchers as they are thought to be more sympathetic, when compared to their male counter-parts (Pollner, 1988). This might be expected to be especially prominent where sensitive topics are discussed - reports of depression, substance abuse, and antisocial behaviours were recorded more often by female interviewers, regardless of the gender of the participant being interviewed (Pollner, 1988). It could be argued that the influence of gender on the willingness to report symptoms to an interviewer could contribute to differences in outcome measurements and some research does suggest that participants may be more open with female researchers (Pollner, 1988). In theory, this could influence measurement scores. However, research into researcher gender differences in RCTs is sparse. It is widely acknowledged that many treatments and interventions have different impacts on male and female participants in terms of biological differences and hence participant gender is a common confounder in RCTs (Phillips & Hamberg, 2016; Siepmann et al., 2016).

Social science research has shown that females in everyday conversation tend to facilitate a more open dialogue and are better at assessing others' feelings and personality traits (Roter & Hall, 2004). There is some evidence in healthcare studies that gender plays a role and some studies have shown that researcher gender has an impact on participant responses particularly in pain studies (Fisher, 2007; McClelland & McCubbin, 2008; Miyazaki & Taylor, 2008). However, the influence of researcher gender is understudied in Dementia trials and no definitive conclusions can be drawn at this stage.

## Research Questions

As a result of the uncertainties around the impact of using multiple researchers to collect QoL outcome data and the potential impact of researcher gender on outcome data collection, the focus of this thesis will explore the influence of these two factors during research into QoL for PwD. The aim of the research is to use data from a previously conducted RCT to determine if there are any differences between the scores of those who had the same researcher during measurement visits compared to those who had different researchers. Two research questions will be explored: in a study with three time points for PwD and carer dyads:

1) Does the use of a different researcher at time points impact upon the outcome measure?
2) Does the (biological) gender of the attending researcher impact upon the outcome measure?

# Chapter Three – Data Description & Exploration

To address the two research questions proposed in Chapter Two, this thesis uses data from the REMCARE study 'Reminiscence groups for people with dementia and their family caregivers – effectiveness and cost-effectiveness pragmatic multi centre randomised trial' which aimed to test joint reminiscence groups for people with mild to moderate Dementia and their family carer run by Robert Woods at Bangor University (Woods et al., 2012). This was a multi-centre parallel two-arm study, which recruited 488 dyads in total (People with Dementia (PwD) and their carer). Dyads were recruited into the study across the UK from several centres including London, Manchester, Bangor, Hull, Bradford and Gwent through several types of mental health services. All PwD recruited were those who met the DSM-IV criteria for Dementia of any type who were in the mild to moderate stage of the disease, living in the community and who had a carer with whom they maintained regular contact with that could participate in the trial.

Those recruited and consented were randomised on a 1:1 allocation ratio to either the treatment arm or the treatment as usual. The participants were stratified by recruitment centre and their relationship to the carer (horizontal or vertical; horizontal being a carer of the same age or generation (e.g. a spouse, friend or sibling) whilst vertical is a different generation (e.g. parent and child relationship)) (Woods et al., 2012). Those randomised to the treatment arm received standard care plus a joint reminiscence-based group therapy and those allocated to the treatment as usual arm received standard care alone (Woods et al., 2012). The main aim of the intervention was to improve the quality of life (QoL) for the person with Dementia (PwD) and to reduce carer-related stress for the caregiver; subsequently improving the participant-carer relationship. On completion, the study showed no evidence for the effectiveness or cost-effectiveness for the intervention (Woods et al., 2012).

Data was collected at baseline, 3 months after baseline (follow-up 1) and 10 months after baseline (follow-up 2) and a variety of outcome measures were collected at each time point covering QoL, memory functionality, depression and anxiety, caregiver's mental health and stress related to caregiving, activities of daily living, service use and the quality of the carer-patient relationship. The REMCARE primary analysis was an ANCOVA model at follow-up 2, with a secondary ANCOVA model run at follow-up 1. Further exploratory general linear mixed models were run at both time points to investigate effects over time. The participants in the study were not blind to group allocation but the researchers collecting the follow-up data were (Woods et al., 2012) however, some were

unintentionally unblinded during interviews. It is not known explicitly how many researchers were unblinded, however, 'perception sheets' completed by researchers indicated that the proportion of correct definite judgements was around 25% at follow-up 2.

The datasets used for this thesis' analysis were created by extracting variables and data from the original REMCARE dataset, merging datasets together and filtering out participants who did not meet necessary criteria, such as those who did not complete all three visits. Two separate datasets containing the corresponding dyads were created and used for analysis, a participant data set (the PwD) and carer dataset (the PwD corresponding carers). All data extraction, merging and subsequent analysis was conducted using SPSS IBM version 25 (IBM, 2017). Information on the researchers collecting data, demographic variables and the required outcome measures were included in each of the datasets. The demographic variables included are age, gender, ethnicity, marital status, who the participant lives with, centre, treatment allocation, wave and date of birth. REMCARE recruitment was conducted in five 'waves' therefore the wave variable is which recruitment wave the participant was recruited and randomised during. The PwD age and PwD gender were also included in the carer dataset as they are required for the carer analysis models to replicate the REMCARE analysis.

PwD completed the assessments with a researcher whilst carers completed their measures independently with little input from the researchers. Due to this, it is hypothesised that a researcher effect, if present, would be seen in the participant data but not in the carer data. Therefore, to explore the research questions, the current analysis requires outcome measures that have been completed by both the participant and the carer, this included the QCPR, QoL-AD and the EQ-5D. The EQ-5D is a relatively short and generic measure of only five items, which can be independently completed by the participants with little interaction from the researcher, thus the measure is unlikely to be susceptible to a researcher effect and therefore is not included for this analysis. The QCPR and QoL-AD are both measures which require interaction and communication amongst the participants and researchers and therefore could be influenced by a researcher. The QCPR was collected from both the PwD and carer based on their own perceptions of the caregiving relationship. The participant QoL-AD was collected from the PwD based on their perception of their own QoL and the carer QoL-AD proxy version was collected from the carer regarding their view of the QoL of the PwD they care for.

As discussed in Chapter Two, the QoL-AD is a common measure used in Dementia studies. The QoL-AD is a validated self-report outcome measure that is used to assess the QoL of a person with Dementia (Logsdon et al., 1999). The measure consists of 13 items scored on a 4-point Likert scale

ranging from 1 – 4 where 1 represents 'poor' and 4 represents 'Excellent'. The final score is calculated by summing the 13 items hence the final score ranges from 13 to 52. A higher total score indicates better QoL. The item has a missing data rule of mean substitution where if two or fewer items are missing then the mean score of the completed items can be used for the missing values. If more than two items are missing, then the total score should not be calculated.

The QCPR is also a validated self-report outcome measure used to assess the relationship between a participant and their carer by evaluating the presence of warmth in their relationship and the absence of conflict and criticism (Spruytte et al., 2002). The measure is less used in Dementia research studies compared to the QoL-AD however, it contains sensitive and personal questions that may be affected by a researcher influence even more so than other measures. The QCPR consists of 14 items where each item is scored on a five-point Likert-scale ranging from 1 to 5. The item options range from "totally not agree" to "totally agree". The measure has two subscales; a conflict scale (scored in reverse) consisting of 6 items and a warmth scale consisting of 8 items. Each subscale is calculated by summing the individual components to give a total score for the subscale. The total score is calculated by summing the two scales; hence the final score can range from 14 to 70 with higher scores indicating a better perceived relationship between participant and caregiver. No missing data rules were found for the measure so if any items are missing the total score cannot be calculated. Appendix 2 summarises the characteristics of the QoL-AD and QCPR.

Along with the researcher data, demographics and the outcome measures, additional variables were created using Microsoft Excel (Microsoft, 2018) for the datasets these include 'number of visits', 'researcher attendance' and 'gender of researcher in attendance'. The 'number of visits' variable corresponds to whether the participant had completed one, two or three visits which was based on whether a variable 'completed by' was answered at each of the time points. As this analysis is exploring a researcher effect across the time points, based on a relationship building hypothesis, it was necessary to filter out those who did not complete all three visits. A high proportion of PwD did complete the three visits (69%), filtering on this resulted in a dataset containing 336 participants. Six participants were removed from both data sets due to the carers not completing all three follow-ups. Therefore, the final data sets contained 330 participants in each totalling 660 (330 dyads - the PwD and their corresponding carer), detailed in Figure 3. All analysis were run using these datasets however as some demographic and outcome data were missing and complete case analysis is conducted (described in more detail later in this Chapter) some of the models did not contain the full sample. The analysis and figures presented will detail the number of cases used. The REMCARE data final sample contained 350 participants however they included data where those who just completed

baseline and one of the follow-ups however this analysis requires participants and carers that completed all three visits.

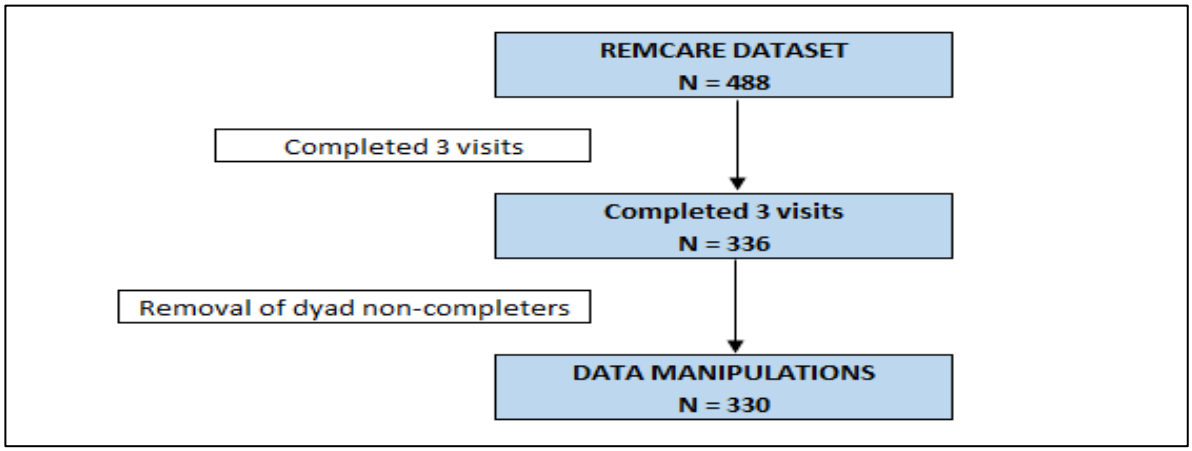


***Figure 3:*** *Flowchart of Number of Cases in Dataset*

The 'researcher attendance' variable was created by categorising participants into groups based on whether they had the same researcher for all visits, different researchers or a combination. At follow-up 1 only two visits had been completed and at follow-up 2 three visits had been completed, therefore two variables were needed - one for the researcher attendance at follow-up 1 (same or different) and one at follow-up 2 (same, different or a combination (as described below)). The variable was coded based on the 'completed by' variable which contained the identifier of the researchers who conducted interviews. For the researcher attendance at follow-up 1, participants were coded into one of two categories '1' indicated that the participant had the same researcher collecting baseline and follow-up 1 data (i.e. "same researcher") and '2' meant that they had a different researcher collecting baseline data to the follow-up 1 data (i.e. "different researchers").

The researcher attendance variable at follow-up 2 was coded in a similar way but participants were put into one of three categories; **one** researcher attending all three visits to collect data (i.e. "same researcher"), **two** visits with the same researcher but one of the visits was completed with a different researcher to the other two visits (i.e. "one same and two different researchers") or lastly, **three** different researchers attending the visits therefore had all different researchers collecting outcome data (i.e. "different researchers"). During the rest of the thesis when referring to 'researcher attendance' or for any statistics and analysis presented, this will be indicating the corresponding researcher attendance variable at either follow-up 1 or follow-up 2.

To conduct secondary exploratory analysis a 'gender of researcher in attendance' variable was created, again using the researcher identifiers in the 'completed by' variable. The variable indicates the whether the participant had the same or different researcher and if different whether the researcher gender was consistent or not. As we are interested in the consistency of the gender of the researcher there is no need to distinguish at follow-up 2 whether the patient had two or three researchers, therefore the same variable was created for follow-up 1 and follow-up 2. Participants were put into one of three categories; the same researcher attended visits therefore same gender, different researchers attended visits but were the same gender, and lastly, different researchers attended visits and were different genders. These variables and categories are all listed in Figure 4.

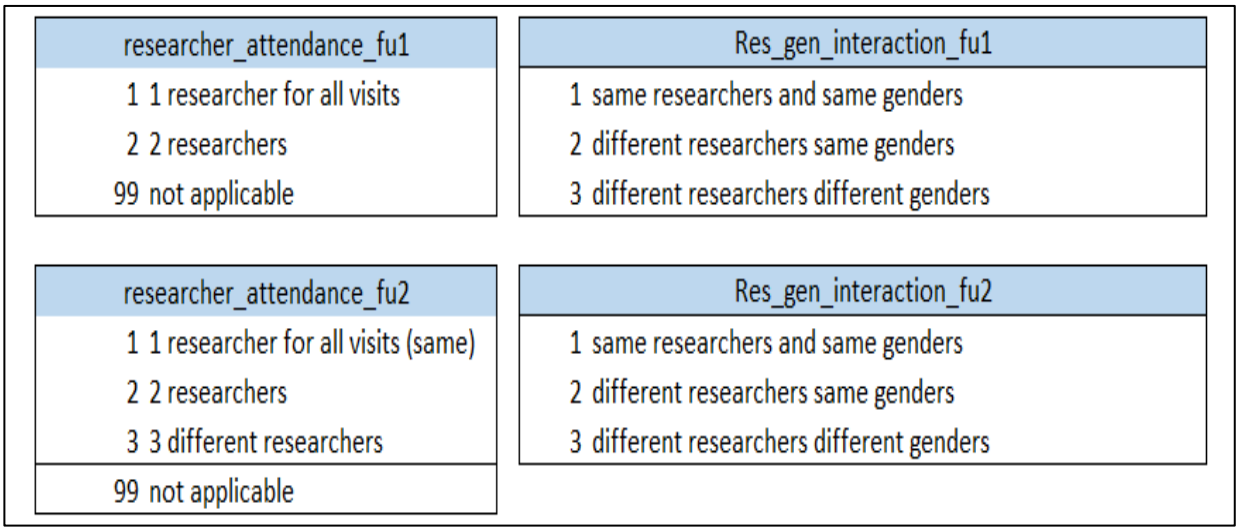


***Figure 4:*** *Coding of Researcher Attendance and Gender of Researcher in Attendance Variables*

Several assumptions were made on researcher names during coding. Firstly, a first initial and last name matching a full name were assumed to be the same researcher for example 'Rachel Evans' and 'R Evans'. Secondly, any minor spelling inconsistencies such as 'Rachael Evans and Rachel Evans' were classed as the same researcher. Additionally, abbreviations of names were categorised as the same researcher for example 'Rach Evans' and 'Rachel Evans'. One other assumption that was made was that if two researchers were listed the first researcher name was taken as the main researcher, this only occurred on one occasion out of 990 possibilities. For some researchers it was difficult to establish the gender based on the name alone, therefore, if the gender was unclear a website ("GenderChecker.com", n.d) giving probabilities of the gender based on the name given was used with the gender assumed by the highest probability given. This could not be done for all as some researchers only initials were given, some were unisex names and one name was not recognisable on the website therefore these researcher genders were left unknown, this was only the case for four researchers.

Table 1 details the frequencies of the researcher attendance and the researcher gender attendance variables at each follow-up. At follow-up 1 the splits between the two attendance groups are almost even with 48% in the same researcher group and 52% in the different researchers group. At follow-up 2 the splits between the three groups are still relatively even with adequate samples representing each group. The gender of the researcher in attendance group splits vary a little more. At follow-up 1 those who had the same researcher and same gender is 160 (49%) those who had different researchers of the same gender is 119 (36%) and those who had different researchers with different genders is 47 (14%). At follow-up 2 the splits across groups vary slightly but still have adequate representation in each group with the majority having different researchers of the same gender (42%), followed by a large proportion having the same researcher therefore the same gender (36%) and slightly less having different researchers with representatives of both genders (21%).

***Table 1:*** *Occurrence of Researcher Attendance at Follow-up 1 and Follow-up 2*

| **Researcher attendance N = 330** | | |
|---|---|---|
| | **Follow-up 1 N (%)** | **Follow-up 2 N (%)** |
| **Same researcher** | 160 (48%) | 118 (36%) |
| **Two different researchers** | 170 (52%) | 129 (40%) |
| **Three different researchers** | *N/A | 83 (25%) |
| **Researcher gender attendance N = 330** | | |
| | **Follow-up 1 N (%)** | **Follow-up 2 N (%)** |
| **Same Researcher (therefore same gender for all visits)** | 160 (49%) | 118 (36%) |
| **Different Researchers for visits of the same genders** | 119 (36%) | 139 (42%) |
| **Different Researchers for visits with representatives of both genders** | 47 (14%) | 69 (21%) |
| **Unknown** | 4 (1%) | 4 (1%) |

*N/A indicates not applicable as at follow-up 1 only two visits would have been carried out

The original data collected was entered at North Wales Organisation for Randomised Trials in Health (NWORTH) via Teleform scanning and subsequent SDV checking and data cleaning were conducted, therefore due to the confidence in the data entry and cleaning process no additional cleaning was required.

Table 2 contains figures of the completeness of the outcome measure total scores. Complete cases are defined as the outcome measures which have enough items answered for the total score to be calculated. Some measures may have been partially completed that will not contribute to the total scores. For the QCPR no missing data rules were found in publications, therefore all items must be

complete for a total score to be calculated. For the QoL-AD the missing rule, as noted in Appendix 2, is that if up to two items are missing then the items can be substituted using the mean of the available items. Appendix 3 details the number of items missing according to the raw scores at each time point. Using the missing data rule, 43 participants and 18 carers at baseline, 59 participants and 19 carers at follow-up 1 and 54 participants and 29 carers at follow-up 2 required their total scores to be calculated using missing items imputation for the QoL-AD measure.

***Table 2:*** *Completion Rates of the Outcome Measures*

| Outcome Measures | Overall N = 330 N (%) | Researcher attendance at fu1 | | Researcher attendance at fu2 | | |
|---|---|---|---|---|---|---|
| | | ***Same researcher N = 160*** N (%) | ***Two different N = 170*** N (%) | ***Same researcher N = 118*** N (%) | ***Two different N = 129*** N (%) | ***Three different N = 82*** N (%) |
| **Participant - Qol-AD total score** | | | | | | |
| ***Baseline*** | **317 (96%)** | 150 (94%) | 167 (98%) | 109 (92%) | 126 (98%) | 82 (100%) |
| ***Follow-up 1*** | **309 (94%)** | 145 (91%) | 164 (97%) | 106 (90%) | 123 (95%) | 80 (96%) |
| ***Follow-up 2*** | **292 (89%)** | N/A | N/A | 99 (84%) | 118 (92%) | 75 (90%) |
| **Participant - QCPR total score** | | | | | | |
| ***Baseline*** | **308 (93%)** | 149 (93%) | 159 (94%) | 108 (92%) | 124 (96%) | 76 (92%) |
| ***Follow-up 1*** | **293 (89%)** | 138 (86%) | 155 (91%) | 103 (87%) | 117 (91%) | 73 (88%) |
| ***Follow-up 2*** | **288 (87%)** | N/A | N/A | 98 (83%) | 114 (88%) | 76 (92%) |
| **Carer - Qol-AD proxy total score** | | | | | | |
| ***Baseline*** | **326 (99%)** | 157 (98%) | 169 (99%) | 115 (98%) | 129 (100%) | 82 (100%) |
| ***Follow-up 1*** | **323 (98%)** | 157 (98%) | 166 (98%) | 116 (98%) | 128 (99%) | 79 (96%) |
| ***Follow-up 2*** | **329 (<100%)** | N/A | N/A | 118 (100%) | 129 (100%) | 82 (100%) |
| **Carer - QCPR total score** | | | | | | |
| ***Baseline*** | **321 (97%)** | 156 (98%) | 165 (97%) | 115 (98%) | 128 (99%) | 78 (94%) |
| ***Follow-up 1*** | **312 (95%)** | 149 (93%) | 163 (96%) | 111 (94%) | 124 (96%) | 77 (93%) |
| ***Follow-up 2*** | **321 (97%)** | N/A | N/A | 117 (99%) | 126 (98%) | 78 (94%) |

Overall, the completion rates of the outcome measures are high and only vary by a small percentage across time points and always remain above 87%. When split by researcher attendance variables the completion rates vary slightly in each group but still remain over 79%. A visual inspection of the completion rate reveals that there appears to be no major differences between those participants who had the same or different researchers across time points, raising no major concerns for analysis. In general, the carers had slightly higher completion rates than the participants across the study, which is to be expected and is usually the case in trials of this nature. There was little variation

between the completion of the outcome measures, however overall the QoL-AD had higher completion rates for the total scores which could be attributed to the missing data rule.

Methods of dealing with missing data in RCTs is debated amongst researchers, although it is acknowledged that methods of dealing with missing data for analysis should be based on how much data is missing, the kind of missing data (single items, full measures, a measurement time point) and what type of missing data exists within the dataset (missing completely at random (MCAR), missing at random (MAR) or missing not at random (MNAR)). The REMCARE study adopted multiple imputation techniques for analysis. Since the completion rates of the outcomes to be used for the current analysis are high, the missing data should not affect the results of the current analysis. Predictors of "missingness" were considered, however, given such little missing data this was deemed not necessary and therefore complete case analysis will be conducted with no methods of missing data imputation adopted.

Table 3 contains the descriptive statistics for both datasets presented overall and split by researcher attendance. Initially, ethnicity and marital status were collected with 22 and 7 categories respectively, however, as there was very little representation in many of the categories, these two variables have been dichotomised into two groups, as the REMCARE study did. Ethnicity has been dichotomised into 'White' and 'other' and marital status into 'Spousal' or 'Non-Spousal'.

A large majority of the sample are White (PwD; 97%, carers; 96%), married (PwD; 75%, carers; 88%) and live with their spouse (PwD; 70%, carers; 79%). The prevalence of males and females in the data is even with an almost equal split for participant (Male; 51%, Female; 49%) but varies more for carers with a higher number of females in the sample (Male; 33%, Female; 67%). Although the representation of samples within these variables is uneven, with low participant samples representing some categories, surprisingly, the proportions across the researcher groups (same or different researcher(s)) for all appear to be evenly split.

Overall, the number of participants at centres are generally even except in the case of Gwent who recruited just 4% of the participants. There appears to be an imbalance between the researcher attendance groups for the centres therefore it is important to consider this when interpreting the analysis results. There is a high percentage of participants in wave one, two and three but fewer in wave four or five. The splits between researcher attendances at both follow-ups are relatively unbalanced in some groups which also needs to be considered when interpreting the analysis results.

The split between allocation group overall has remained relatively even and is well balanced across the researcher attendance groups.

On average, the age of the PwD was approximately 77 years old and for the carers the average age was approximately 69 years old. The age range of the carers is larger than the age range of the participants (23 – 90 and 54 – 93 respectively). The mean ages between the researcher attendance groups (at both follow-ups) are relatively similar for participants and carers.

There is a small amount of missing observations of demographic data, which will have a minor impact on the number of participants used in the analysis models. In total there are 6 missing observations for the participant demographic data and 12 for the carer data, some of these missing cases are non-independent with one participant being responsible for three of these observations and several carers accounting for several observations. Therefore, with no missing data imputation methods being adopted this will result in 3 participants and 5 carers being excluded from the analysis models. Therefore, a further eight dyads are removed and a maximum sample of 322 participants will be used in the full analysis models, indicated in Figure 5.

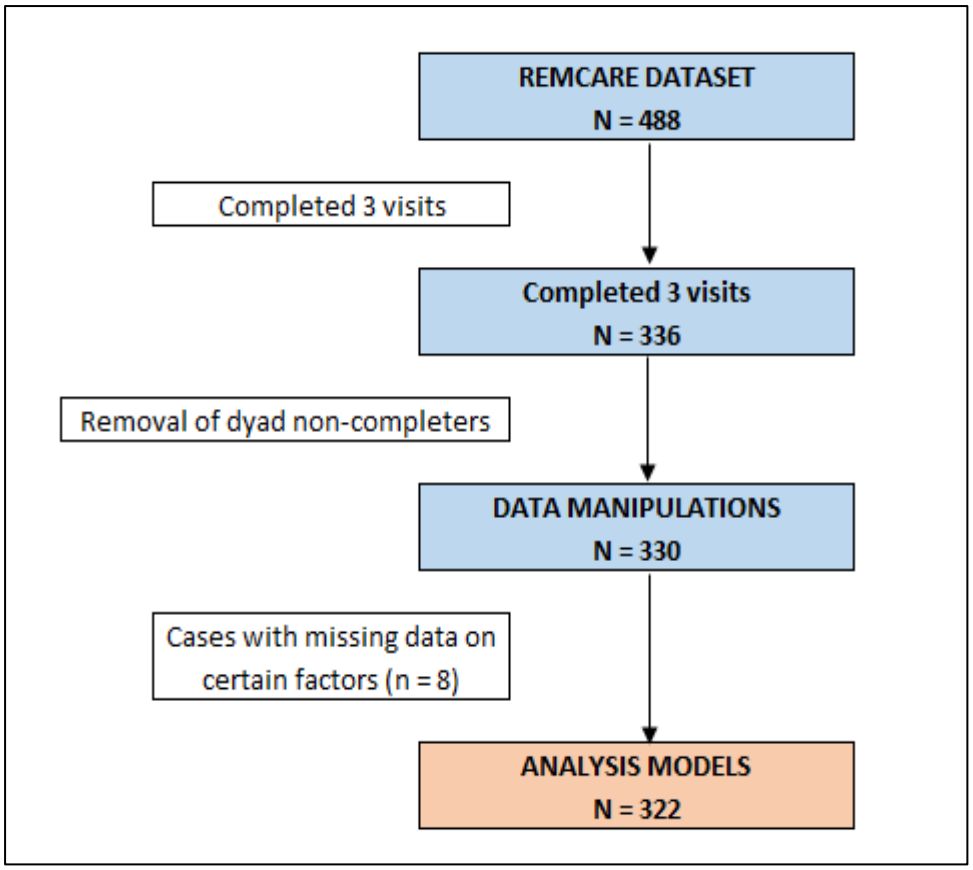


***Figure 5:*** *Flowchart of Number of Cases for Analysis Models*

***Table 3:*** *Descriptive Statistics of Demographics and Other Characteristics*

| Data set | Variable | | Overall N = 330 N (%) | Researcher Attendance at follow-up 1 | | Researcher attendance at follow-up 2 | | |
|---|---|---|---|---|---|---|---|---|
| | | | | ***Same researcher N = 160*** N (%) | ***Two different N = 170*** N (%) | ***Same researcher N = 118*** N (%) | ***Two different N = 129*** N (%) | ***Three different N = 83*** N (%) |
| **Participant** | **Gender** | **Female** | **162 (49%)** | 82 (51%) | 80 (47%) | 61 (52%) | 63 (49%) | 38 (46%) |
| | | **Male** | **167 (51%)** | 78 (49%) | 89 (52%) | 57 (48%) | 66 (51%) | 44 (53%) |
| | | **Missing** | **1 (<1%)** | 0 | 1 (<1%) | 0 | 0 | 1 (1%) |
| | **Ethnicity** | **White** | **319 (97%)** | 157 (98%) | 162 (95%) | 115 (98%) | 127 (98%) | 77 (93%) |
| | | **Other** | **10 (3%)** | 2 (1%) | 8 (5%) | 2 (2%) | 2 (2%) | 6 (7%) |
| | | **Missing** | **1 (<1%)** | 1 (<1%) | 0 | 1 (<1%) | 0 | 0 |
| | **Marital status** | **Spousal** | **246 (75%)** | 121 (76%) | 125 (74%) | 86 (73%) | 98 (76%) | 62 (75%) |
| | | **Non-spousal** | **82 (25%)** | 38 (24%) | 44 (26%) | 31 (26%) | 30 (23%) | 21 (25%) |
| | | **Missing** | **2 (<1%)** | 1 (<1%) | 1 (<1%) | 1 (<1%) | 1 (<1%) | 0 |
| | **Lives with** | **No-one** | **49 (15%)** | 26 (16%) | 23 (14%) | 19 (16%) | 18 (14%) | 12 (15%) |
| | | **Other** | **13 (4%)** | 3 (2%) | 10 (6%) | 2 (2%) | 8 (6%) | 3 (4%) |
| | | **Other family** | **26 (8%)** | 14 (9%) | 12 (7%) | 13 (11%) | 7 (5%) | 6 (7%) |
| | | **Spouse** | **229 (70%)** | 111 (69%) | 118 (69%) | 80 (68%) | 90 (70%) | 59 (71%) |
| | | **Spouse & other family** | **11 (3%)** | 4 (3%) | 7 (4%) | 2 (2%) | 6 (5%) | 3 (4%) |
| | | **Missing** | **2 (<1%)** | 2 (1%) | 0 | 2 (2%) | 0 | 0 |
| **Carer** | **Gender** | **Female** | **221 (67%)** | 107 (67%) | 114 (67%) | 80 (68%) | 84 (65%) | 57 (69%) |
| | | **Male** | **108 (33%)** | 53 (33%) | 55 (32%) | 38 (32%) | 45 (35%) | 25 (30%) |
| | | **Missing** | **1 (<1%)** | 0 | 1 (<1%) | 0 | 0 | 1 (1%) |
| | **Ethnicity** | **White** | **317 (96%)** | 115 (97%) | 162 (95%) | 113 (96%) | 126 (98%) | 78 (94%) |
| | | **Other** | **10 (3%)** | 3 (2%) | 7 (4%) | 3 (3%) | 3 (2%) | 4 (5%) |
| | | **Missing** | **3 (1%)** | 2 (1%) | 1 (<1%) | 2 (1%) | 0 | 1 (1%) |
| | **Marital status** | **Spousal** | **290 (88%)** | 142 (89%) | 148 (87%) | 108 (92%) | 108 (84%) | 74 (89%) |
| | | **Non-spousal** | **36 (11%)** | 16 (10%) | 20 (12%) | 9 (8%) | 19 (15%) | 8 (10%) |
| | | **Missing** | **4 (1%)** | 2 (1%) | 2 (1%) | 1 (<1%) | 2 (1%) | 1 (1%) |
| | **Lives with** | **No-one** | **10 (3%)** | 6 (4%) | 4 (2%) | 2 (2%) | 6 (5%) | 2 (2%) |
| | | **Other** | **14 (4%)** | 5 (3%) | 9 (5%) | 4 (3%) | 8 (6%) | 2 (2%) |
| | | **Other family** | **26 (8%)** | 15 (9%) | 11 (7%) | 11 (9%) | 9 (7%) | 6 (7%) |
| | | **Spouse** | **260 (79%)** | 127 (79%) | 113 (78%) | 95 (81%) | 99 (77%) | 66 (80%) |
| | | **Spouse & other family** | **15 (5%)** | 4 (3%) | 11 (7%) | 4 (3%) | 6 (5%) | 5 (6%) |
| | | **Other & no one** | **1 (<1%)** | 0 | 1 (<1%) | 0 (0) | 0 (0) | 1 (1%) |
| | | **Missing** | **4 (1%)** | 3 (2%) | 1 (<1%) | 2 (2%) | 1 (1%) | 1 (1%) |

| Data set | Variable | | Overall N = 330 N (%) | Researcher Attendance at follow-up 1 | | Researcher attendance at follow-up 2 | | |
|---|---|---|---|---|---|---|---|---|
| | | | | *Same researcher* ***N = 160*** N (%) | *Two different* ***N = 170*** N (%) | *Same researcher* ***N = 118*** N (%) | *Two different* ***N = 129*** N (%) | *Three different* ***N = 83*** N (%) |
| **BOTH CARER AND PARTICIPANT** | **Centre** | **Bangor** | **57 (17%)** | 41 (26%) | 16 (9%) | 39 (33%) | 18 (14%) | 0 |
| | | **Bradford** | **39 (12%)** | 35 (22%) | 4 (2%) | 27 (23%) | 12 (9%) | 0 |
| | | **Gwent** | **14 (4%)** | 8 (5%) | 6 (4%) | 3 (3%) | 8 (6%) | 3 (4%) |
| | | **Hull** | **46 (14%)** | 11 (7%) | 35 (21%) | 11 (9%) | 24 (19%) | 11 (13%) |
| | | **London North** | **59 (18%)** | 13 (8%) | 46 (27%) | 5 (4%) | 15 (12%) | 39 (47%) |
| | | **London South** | **49 15%)** | 14 (9%) | 35 (21%) | 8 (7%) | 16 (12%) | 25 (30%) |
| | | **Manchester** | **66 (20%)** | 38 (24%) | 28 (16%) | 25 (21%) | 36 (28%) | 5 (6%) |
| | **Wave** | **Wave 1** | **101 (31%)** | 55 (34%) | 46 (27%) | 44 (37%) | 35 (27%) | 22 (27%) |
| | | **Wave 2** | **82 (25%)** | 37 (23%) | 45 (26%) | 27 (23%) | 24 (19%) | 31 (37%) |
| | | **Wave 3** | **88 (27%)** | 42 (26%) | 46 (27%) | 29 (25%) | 44 (34%) | 15 (18%) |
| | | **Wave 4** | **54 (16%)** | 26 (16%) | 28 (17%) | 18 (15%) | 22 (17%) | 14 (17%) |
| | | **Wave 5** | **5 (2%)** | 0 | 5 (3%) | 0 | 4 (3%) | 1 (1%) |
| | **Allocation** | **Intervention - Group 1** | **192 (58%)** | 94 (59%) | 98 (58%) | 68 (58%) | 71 (55%) | 53 (64%) |
| | | **Control - Group 2** | **138 (42%)** | 66 (41%) | 72 (42%) | 50 (42%) | 58 (45%) | 30 (36%) |

| Data set | Variable | Overall N Mean (SD) Range | Researcher attendance at follow-up 1 | | Researcher attendance at follow-up 2 | | |
|---|---|---|---|---|---|---|---|
| | | | *Same Researcher* N Mean (SD) Range | *Two Different* N Mean (SD) Range | *Same researcher* N Mean (SD) Range | *Two different* N Mean (SD) Range | *Three different* N Mean (SD) Range |
| **Participant** | **Age** | **330** | 160 | 170 | 118 | 129 | 83 |
| | | **77 (7.60)** | 77 (7.86) | 78 (7.31) | 77 (7.94) | 77 (7.61) | 79 (6.93) |
| | | **54 – 93** | 54 - 91 | 56 - 93 | 54 - 91 | 56 - 93 | 62 - 93 |
| **Carer** | | **330** | 160 | 170 | 118 | 129 | 83 |
| | | **69 (11.38)** | 68 (11.85) | 71 (10.80) | 68 (12.20) | 69 (11.45) | 72 (9.68) |
| | | **23 - 90** | 36 - 89 | 23 - 90 | 36 - 89 | 23 - 87 | 43 - 90 |

Basic statistical tests, as outlined below, were conducted on each of the factor variables to investigate if there are any individual effects on the outcome measures or correlations amongst variables. The exploratory tests have been conducted on all of the factors used in the REMCARE final analysis models and the two additional variables created for this research (the researcher attendance and gender of researcher in attendance variables). These individual exploratory tests do not take into account any baseline differences or any other of the factor effects but will give further insight of any individual variable effects. Ethnicity and 'Lives with' were not used in the final REMCARE model, therefore will not be included in the current research. A large majority of participants' ethnicity was recorded as white and only small samples represented the other categories therefore the variable would not have led to any robust results. Lives with was proxy represented by the horizontal/vertical carer relationship which was a stratification variable and therefore did not need to be additionally included in the models.

For the two-level categorical variables independent t-tests were conducted and for variables with more than two categories the equivalent ANOVA tests were run to investigate any differences between the mean outcome measure scores amongst the variable levels. The assumption checks that are the same for both models (a full list of these can be found in Appendix 4) were conducted and have all been evaluated to sufficiently hold to run the models. Box plots identified some observations as outliers, however the outliers identified are not large and are within the expected range of the scales and therefore are not excluded nor do they require any transformation. Some factors did not satisfy the assumption of homogeneity; the results reported for those that satisfied the assumption are the assumed equal variance results and for those that did not meet the assumption are the 'assumed unequal variance' for the t-tests or the Welch test results for ANOVAs.

The results of the tests are displayed in Appendix 5. The researcher attendance variable produced no statistically significant differences on the QCPR or QoL-AD at either follow-up for the PwD or carer data, indicated in Table 1 in Appendix 5. The gender of researcher in attendance, results of which are also contained in Table 1 in Appendix 5, is significant at the 5% level at follow-up 1 on the PwD QCPR $F(2, 286) = 3.28$, $p = 0.04$, on the carer QCPR $F(2, 305) = 2.96$, $p = 0.05$, and at the 1% level of significance on the carer proxy QoL-AD $F(2, 316) = 4.40$, $p = 0.01$. At follow-up 2 the gender of researcher in attendance is not statistically significant at the 5% level on any of the outcome measures.

The results of the t-tests conducted on gender, marital status and allocation are contained in Table 2 of Appendix 5. There is a statistically significant difference between PwD gender groups at the 5% level on the PwD QoL-AD measure at follow-up 1, $t(309) = -4.17$, $p = 0.04$, with a mean difference

of 1.32 (95% CI; 0.04 to 2.59) with Females having overall higher QoL scores. There are no other statistically significant differences of the PwD gender on the PwD measures.

Statistically significant differences were noted on the carer outcomes at the 1% level between the carer gender groups with males on average having higher scores. On the carer QCPR at follow-up 1 a mean difference of -4.33 was noted ($t(309) = -4.17$, $p < 0.01$) and at follow-up 2 a mean difference of -3.65 was found ($t(318) = -3.27$, $p < 0.01$). On the carer proxy QoL-AD at follow-up 1 a mean difference of -2.28 was present ($t(320) = -3.13$, $p < 0.01$) and at follow-up 2 again a mean difference of -2.28 ($t(326) = -3.24$, $p < 0.01$). The PwD gender is also statistically significant at the 1% level on the carer QCPR outcomes at follow-up 1, $t(301) = 3.83$, $p < 0.01$, and at follow-up 2, $t(318) = 2.59$, $p = 0.01$, with mean differences of 3.76 and 2.74 with female PwD's on average having higher scores. PwD gender is not statistically significant on the carer proxy QoL-AD outcomes at either time points.

The PwD marital status, carer marital status and group Allocation variables are not statistically significant at either follow-ups on any of the PwD measures or the carer measures. Results of the wave variable ANOVA are detailed in Table 3 of Appendix 5, which also is not significant on any of the outcome measures at both time points.

The results of the ANOVA model run on the centre variable are displayed in Table 4 in Appendix 5. The variable produced a statistically significant result at the 1% level $F(6,285) = 2.96$, $p = 0.01$, on the PwD QoL-AD at follow-up 1, to identify where these differences lie, post hoc tests would need to be conducted. These results should be treated with caution, as there is a variety of representation at each centre. The variable is not significant at the 1% or 5% level on any other of the PwD outcomes or carer outcomes at either follow-up.

To investigate any relationships between continuous factors estimates of Pearson's product-moment correlation were obtained. The assumptions of Pearson's test have been checked, a full list of which can be found in Appendix 4 and have all been evaluated to hold. The Pearson's results, found in Appendix 5 Table 5, indicate a very small (Cohen, 1988) positive correlation between age of participant and QCPR scores at follow-up 2, significant at the 5% level ($p = 0.04$) with a correlation coefficient of $r(286) = 0.12$. This indicates that overall as the PwD's age increases the QCPR scores also increase. There is also a small positive (Cohen, 1988) correlation at the 1% level of significance between age of carer and carer QoL-AD proxy scores at follow-up 1 and between the PwD age and carer proxy QoL-AD scores at follow-up 2 with a correlation coefficient of 0.16 and - 0.14 respectively.

For this measure overall as the carer age increases the scores of the carer proxy QoL increase however, as the PwD age increases the proxy QoL scores decrease. As expected, the outcome measure baseline scores all significantly correlate with the follow-up scores at the 1% level of significance. All other correlations evaluated were not statistically significant at the 5% level.

To assess any associations between the researcher attendance and gender of researcher in attendance variables with other categorical variables Pearson's Chi square tests of independence have been conducted, results of which are detailed in Appendix 5 Table 6. The assumptions of the test have been checked and all hold, the full assumption list and methods to evaluate can be found in Appendix 4. There is a statistically significant result at the 1% level of significance on centre between the researcher variable and the researcher gender attendance variable at follow-up 1 and 2. All four associations are moderately strong according to Cohen's (1988) guidelines for interpreting Cramer's V. The wave variable is statistically significant at the 1% level at follow-up 2 with the researcher attendance variable and is statistically significant with the researcher gender attendance variable at both follow-up 1 and 2 at the 5% and 1% level respectively. Again, results concerning the centre variable should be treated with caution as there are small data samples representing some centres.

**Outcome measures**

Descriptive statistics of the outcome measures at each time point are displayed in Table 4 below, presented overall and split between the researcher attendance groups. The splits between the researcher attendance groups at the mean scores are generally evenly distributed across groups and there appear to be no major differences between any levels of the variable only varying slightly on all outcome measure totals.

As indicated in Appendix 2 for both measures higher scores indicate a better result in terms of outcome measure results; for the QCPR higher scores indicate a better perceived relationship between the dyads and for the QoL-AD measure higher scores indicate a better quality of life. In general, scores for the participant and the carer for both outcome measures decrease slightly across time points. Most remain similar but with a tiny decrease, from baseline to follow-up 1 and then a little more of a decrease at follow-up 2 but not by a large extent. Overall, the carers have lower scores on both measures compared with the participant scores, as seen in Table 4. This is expected as carers typically score the PwD's QoL lower than the patients themselves and generally perceive the quality of the carer-patient relationship lower than the PwD.

***Table 4:*** *Descriptive Statistics of the Outcome Measures*

| Outcome Measure | Overall | | Researcher attendance at follow-up 1 | | | | Researcher attendance at follow-up 2 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | *Same researcher N = 160* | | *Two different N = 170* | | *Same researcher N = 118* | | *Two different N = 129* | | *Three different N = 83* | |
| | N | Mean (SD) Range | N | Mean (SD) Range | N | Mean (SD) Range | N | Mean (SD) Range | N | Mean (SD) Range | N | Mean (SD) Range |
| **Participant - Qol-AD total** | | | | | | | | | | | | |
| ***Baseline*** | **317** | **37.5 (5.18)** | 150 | 37.3 (5.42) | 167 | 37.7 (4.97) | 109 | 37.2 (5.36) | 126 | 37.5 (5.18) | 82 | 38.0 (4.97) |
| | | **21.0 – 52.0** | | 21.0 – 50.0 | | 24.0 – 52.0 | | 21.0 – 50.0 | | 24.0 – 52.0 | | 28.0 – 50.0 |
| ***Follow-up 1*** | **309** | **37.1 (5.71)** | 145 | 36.8 (6.11) | 164 | 37.2 (5.36) | 106 | 36.4 (6.33) | 123 | 37.6 (5.28) | 80 | 37.0 (5.50) |
| | | **17.3 – 50.0** | | 17.3 – 50.0 | | 22.0 – 50.0 | | 17.33 – 50.0 | | 23.0 – 50.0 | | 22.0 – 50.0 |
| ***Follow-up 2*** | **292** | **36.4 (5.48)** | | N/A | | | 99 | 35.5 (5.48) | 118 | 37.2 (5.05) | 75 | 36.5 (6.03) |
| | | **22.0 – 51.0** | | | | | | 22.0 – 51.0 | | 26.0 – 49.0 | | 22.0 – 50.0 |
| **Participant - QCPR total** | | | | | | | | | | | | |
| ***Baseline*** | **308** | **58.3 (6.04)** | 149 | 57.9 (6.43) | 159 | 58.7 (5.65) | 108 | 57.3 (5.77) | 124 | 60.0 (6.20) | 76 | 58.6 (6.07) |
| | | **38.0 – 70.0** | | 38.0 – 70.0 | | 45.0 – 70.0 | | 45.0 – 70.0 | | 38.0 – 70.0 | | 46.0 – 70.0 |
| ***Follow-up 1*** | **293** | **58.1 (6.62)** | 138 | 58.2 (6.82) | 155 | 58.1 (6.45) | 103 | 57.9 (6.65) | 117 | 58.4 (6.74) | 73 | 58.0 (6.45) |
| | | **38.0 – 70.0** | | 42.0 – 70.0 | | 38.0 – 70.0 | | 42.0 – 70.0 | | 41.0 – 70.0 | | 38.0 – 70.0 |
| ***Follow-up 2*** | **288** | **57.3 (6.47)** | | N/A | | | 98 | 57.4 (6.41) | 114 | 56.6 (6.09) | 76 | 58.2 (7.05) |
| | | **34.0 – 70.0** | | | | | | 39.0 – 69.0 | | 35.0 – 70.0 | | 34.0 – 70.0 |
| **Carer – proxy Qol-AD total** | | | | | | | | | | | | |
| ***Baseline*** | **326** | **32.0 (6.13)** | 157 | 31.8 (6.22) | 169 | 32.1 (6.06) | 115 | 31.0 (6.16) | 129 | 33.0 (5.95) | 82 | 31.7 (6.19) |
| | | **15.0 – 48.0** | | 15.0 – 48.0 | | 20.0 – 48.0 | | 15.0 – 48.0 | | 20.0 – 48.0 | | 20.0 – 45.5 |
| ***Follow-up 1*** | **323** | **31.1 (6.24)** | 157 | 30.9 (6.31) | 166 | 31.2 (6.18) | 116 | 30.4 (6.25) | 128 | 31.9 (6.29) | 79 | 30.8 (6.08) |
| | | **15.0 – 51.0** | | 15.0 – 51.0 | | 18.0 – 48.0 | | 15.0 – 48.0 | | 18.0 – 51.0 | | 20.0 – 48.0 |
| ***Follow-up 2*** | **329** | **30.2 (6.06)** | | N/A | | | 118 | 29.7 (5.90) | 129 | 30.7 (6.40) | 82 | 30.1 (5.72) |
| | | **15.4 – 49.0** | | | | | | 19.0 – 47.0 | | 15.4 – 49.0 | | 16.0 – 43.0 |
| **Carer QCPR total** | | | | | | | | | | | | |
| ***Baseline*** | **321** | **54.3 (8.58)** | 156 | 53.7 (8.88) | 165 | 55.0 (8.27) | 115 | 53.0 (8.75) | 128 | 55.5 (7.97) | 78 | 54.3 (9.13) |
| | | **30.0 – 70.0** | | 32.0 – 70.0 | | 30.0 – 70.0 | | 32.0 – 70.0 | | 31.0 – 70.0 | | 30.0 – 70.0 |
| ***Follow-up 1*** | **312** | **53.6 (8.83)** | 149 | 53.0 (9.29) | 163 | 54.2 (8.37) | 111 | 52.2 (8.94) | 124 | 54.6 (8.41) | 77 | 53.9 (9.18) |
| | | **28.0 – 70.0** | | 31.0 – 70.0 | | 28.0 – 70.0 | | 32.0 – 70.0 | | 30.0 – 70.0 | | 28.0 – 70.0 |
| ***Follow-up 2*** | **321** | **53.1 (9.56)** | | N/A | | | 117 | 52.2 (9.97) | 126 | 53.0 (9.47) | 78 | 54.6 (8.98) |
| | | **21.0 – 70.0** | | | | | | 25.0 – 70.0 | | 21.0 – 70.0 | | 31.0 – 70.0 |

The univariate tests indicate significant differences, associations and correlations on some variables on some of the outcome measures, but not all have significant findings. However, regardless of the univariate tests, all the factors are to be included in the analysis models. This is because when evaluating these effects with basic univariate tests we are only looking at the impact of the factors on the outcome measures individually. Factors as a combination can contribute to effects seen; when a new factor is introduced to a model, the others are affected and some factors may not be significant but may still have an importance in the model and taking the approach of significant at univariate analysis could result in important confounding variables being missed. Statistical univariate analysis alone should not decide what factors are entered into a multivariate model in clinical trials and clinical input or past research should guide what variables are required in the model (Heinze & Dunkler, 2017). The analysis models adopted here, described in more detail in Chapters 4 and 5 will follow the same format of those run in the REMCARE analysis with the addition of the variables of interest being included in the models to be able to assess any impact these additional factors have on the data.

The current Chapter has described in detail the dataset which will be used for analysis to evaluate the two research questions, as stated above Chapters Four and Five will detail the analysis of these two questions and the results of these analyses.

# Chapter Four – Analysis of Researcher Attendance (Research Question 1)

The current chapter will explore the first research question of 'In a study with three time points does the use of a different researcher collecting data at the time points impact upon the outcome measure?' To assess this, the aim of the analysis is to explore whether there is an impact on the outcome measure scores (QoL-AD and QCPR measures) based on whether a participant had the same researcher, different researchers or a combination of same and different. Analysis of any statistical differences between follow-up scores of the researcher attendance levels will be conducted and any differences identified will be explored. Analysing whether there is a difference between the groups will give an indication whether or not there is a potential presence of a researcher effect.

## 4.1 Analysis methods

An ANCOVA model was used to evaluate whether there are any statistically significant differences between the different levels of researcher attendance reflected in the outcome measure, whilst taking into account other factors which might have an impact on the scores. In total, eight separate models were run- four on the participant data and four on the carer data at the two follow-ups for both QoL-AD and QCPR outcome measures; Figure 6 lists all analysis models conducted to explore the research question.

Participant Dataset

1. QCPR total score follow-up 1
2. QoL-AD total score follow-up 1
3. QCPR total score follow-up 2
4. QoL-AD total score follow-up 2

Carer Dataset

1. QCPR total score follow-up 1
2. QoL-AD proxy total score follow-up 1
3. QCPR total score follow-up 2
4. QoL-AD proxy total score follow-up 2

***Figure 6:*** *List of Dependent Variables in the Primary Analysis Models*

The models used corresponded with the models that were run for the REMCARE study analysis. The outcome measures were entered as the dependent variables with the corresponding baseline scores included as covariates to adjust for any differences amongst the groups at baseline. Other factors that were entered into the model were entered as fixed effect factors if the variables were categorical and as covariates if they were continuous variables. The REMCARE analysis included all factors as fixed effects apart from Centre and Wave, which were entered in the models as random effects. For the current analysis models all factors have been entered as fixed effects including Centre and Wave. The Wave variable in REMCARE was based on what recruitment period the participants were randomised during therefore is an ambiguous factor and in this data can be treated as fixed.

Similarly, Centre in REMCARE was treated as random as the analysis set was a sample of the population but as we are just concerned with the current sample and can treat the variable as fixed here.

For the participant analysis the variables included as factors in the model were - PwD age, PwD gender, PwD marital status, centre, wave, allocation, the interaction between centre and allocation, the researcher attendance variable (at the corresponding time point; follow-up 1 or 2) and also the interaction between centre and researcher attendance. The carer models had the same factors with the addition of the PwD age and PwD gender as in the REMCARE study.

All analyses were run as complete case analysis without using any missing data imputation techniques as discussed in Chapter Three, due to the way the participant sample was created for this study, the completion rates for outcomes are high and missing data should not impact the results. The assumptions associated with an ANCOVA model have been checked for all of the models; a full list of the assumptions and methods to check can be found in Appendix 6.

## 4.2 Analysis Results

### Assumption checks

Visual inspection of scatterplots revealed that there was a linear relationship between each of the covariates and dependent variable on each level of the researcher attendance group for all models. The assumption of homogeneity of regression slopes was violated in several cases, where this assumption was violated as recommended by Grace-Martin (2013) the interaction term was included in the final model. The list of the interaction terms are contained in Appendix 7 to indicate which models satisfied the assumption and which did not, the interaction terms that were significant and hence included in the main models are reported within the analysis results table.

There were no substantial outliers to consider. Any potential outliers observed in the data were all within expected range of the measures and therefore not removed from the analysis. The deviations were not substantial enough to require consideration of transformations of the data. The Q-Q plots of the residuals indicated that a “perfect” normal distribution was not present; however, the data is only slightly skewed at the tails, therefore, transformations to the data are not required and the assumption *sufficiently* holds. Lastly, the assumptions of homoscedasticity and homogeneity of variances were met for all models.

### 4.2.1 Participant (PwD) data results

Table 5 details the results of the ANCOVA models run on the participant data and Table 6 details the estimated marginal means for the researcher attendance variable having adjusted for the covariates in the models. Where significance was indicated then the associated effect sizes and confidence intervals are presented.

***Table 5**: ANCOVA Model Results for PwD Data*

| QCPR follow-up 1 | | | | QoL-AD follow-up 1 | | | |
|---|---|---|---|---|---|---|---|
| **Factor** | **DF** | **F-value** | **p-value** | **Factor** | **DF** | **F-value** | **p-value** |
| **QCPR Baseline** | 1 | 104.5 | **<0.01 | **QoLAD Baseline** | 1 | 215.22 | **<0.01 |
| **Age** | 1 | 0.05 | 0.83 | **Age** | 1 | 1.28 | 0.26 |
| **Gender** | 1 | 0.74 | 0.40 | **Gender** | 1 | 2.59 | 0.11 |
| **Marital status** | 1 | 0.71 | 0.40 | **Marital status** | 1 | 0.02 | 0.89 |
| **Centre** | 6 | 0.34 | 0.92 | **Centre** | 6 | 1.78 | 0.10 |
| **Wave** | 4 | 2.27 | 0.06 | **Wave** | 4 | 0.49 | 0.74 |
| **Allocation** | 1 | 0.17 | 0.68 | **Allocation** | 1 | 0.92 | 0.34 |
| **Centre x Allocation** | 6 | 0.19 | 0.98 | **Centre x Allocation** | 6 | 2.22 | *0.04 |
| **Researcher Attendance** | 1 | 5.65 | *0.02 | **Researcher Attendance** | 1 | 10.24 | **<0.01 |
| **Centre x Researcher Attendance** | 6 | 1.35 | 0.23 | **Centre x Researcher Attendance** | 6 | 0.89 | 0.50 |
| **QCPR B x Researcher Attendance** | 1 | 6.00 | *0.02 | **QoLAD B x Researcher Attendance** | 1 | 9.42 | **<0.01 |
| **Error (SS within)** | 246 | | | **Error (SS within)** | 267 | | |
| **QCPR follow-up 2** | | | | **QoL-AD follow-up 2** | | | |
| **Factor** | **DF** | **F-value** | **p-value** | **Factor** | **DF** | **F-value** | **p-value** |
| **QCPR Baseline** | 1 | 43.28 | **<0.01 | **QoLAD Baseline** | 1 | 88.94 | **<0.01 |
| **Age** | 1 | 6.01 | *0.02 | **Age** | 1 | 0.45 | 0.50 |
| **Gender** | 1 | 0.00 | 0.99 | **Gender** | 1 | 0.29 | 0.60 |
| **Marital status** | 1 | 0.43 | 0.51 | **Marital status** | 1 | 0.86 | 0.36 |
| **Centre** | 6 | 0.85 | 0.53 | **Centre** | 6 | 2.50 | *0.02 |
| **Wave** | 4 | 0.53 | 0.71 | **Wave** | 4 | 1.14 | 0.34 |
| **Allocation** | 1 | 0.50 | 0.48 | **Allocation** | 1 | 0.27 | 0.60 |
| **Centre x Allocation** | 6 | 0.68 | 0.67 | **Centre x Allocation** | 6 | 0.83 | 0.55 |
| **Researcher Attendance** | 2 | 2.93 | 0.06 | **Researcher Attendance** | 2 | 1.14 | 0.32 |
| **Centre x Researcher Attendance** | 10 | 2.16 | *0.02 | **Centre x Researcher Attendance** | 10 | 0.74 | 0.69 |
| **QCPR B x Researcher Attendance** | 2 | 3.15 | *0.05 | **Error (SS within)** | 246 | | |
| **Error (SS within)** | 234 | | | | | | |

**Significant at the 0.05 level. **Significant at the 0.01 level.*

***Table 6:*** *Adjusted Means for Researcher Attendance Groups from PwD ANCOVA Model*

| OUTOME MEASURE | ADJUSTED VALUES AT FOLLOW-UP 1 | | | | ADJUSTED MEAN DIFFERENCE (P – VALUE) | EFFECT SIZE (95% CI) |
|---|---|---|---|---|---|---|
| | Same Researcher | | Two Different Researchers | | | |
| | N | Mean (SE) | N | Mean (SE) | | |
| **QCPR** | 132 | 58.0 (0.96) | 144 | 57.9 (1.01) | 0.07 (p = 0.95) | -0.01 (-0.25, 0.23) |
| **QOL-AD** | 136 | 37.2 (0.64) | 161 | 37.8 (0.72) | -0.66 (p = 0.36) | 0.074 (-0.16, 0.30) |

| OUTOME MEASURE | ADJUSTED VALUES AT FOLLOW-UP 2 | | | | | |
|---|---|---|---|---|---|---|
| | Same Researcher | | Two Different Researchers | | Three Different Researchers | |
| | N | Mean (SE) | N | Mean (SE) | N | Mean (SE) |
| **QCPR** | 92 | 58.0 (1.10) | 108 | 57.3 (0.86) | 70 | 57.8 (1.34) |
| **QOL-AD** | 91 | 36.7 (0.87) | 115 | 38.0 (0.68) | 74 | 37.5 (0.97) |

QOL-AD at follow-up 1

Detailed in Table 5, the researcher attendance variable is significant at the 1% level on the QoL-AD outcome at follow-up 1, $F(1, 267) = 10.24$, $p < 0.01$. An adjusted mean difference between the researcher groups, presented in Table 6, of -0.66, was found, with those in the same researcher group having an overall lower mean than those in the different researcher group. The difference obtained is not statistically significant at the 5% level ($p = 0.36$) and produced a Cohen's *d* effect size of 0.07 which is classed as a very small effect size (Cohen, 1988). This indicates that the difference is minor and potentially clinically insignificant (Walker, 2008). However, as the interaction term between the independent variable and a covariate was included in the model, this finding should be explored/interpreted in relation to this interaction; this is detailed further in the 'Post hoc tests and further investigation of significant findings' section below.

The only other variables found to be statistically significant at the 5% level on the QoL-AD at follow-up 1 is the interaction between centre and allocation $F(6, 267) = 2.22$, $p = 0.04$. Post hoc tests would need to be carried out to evaluate this effect.

QCPR at follow-up 1

On the QCPR outcome at follow-up 1 there is a statistically significant result on the researcher attendance variable at the 5% level $F(1, 246) = 5.65$, $p = 0.02$, given in Table 5. The estimated adjusted

mean difference is 0.07, shown in Table 6, with those who have the same researcher throughout having a slightly higher overall mean than those in the different researcher groups. This difference is not statistically significant at the 1% or 5% level (p = 0.95) and has a Cohens d effect size of -0.01 which is a very small effect size (Cohen, 1988). Again, this indicates that although the variable in the model is statistically significant the mean difference between the groups is trivial. The effect of the variable requires further inspection and again is detailed in the 'Post hoc tests and further investigation of significant findings' section below.

### QOL-AD at follow-up 2

At follow-up 2 the researcher attendance variable is not significant at the 5% level On the QoL-AD $F(2, 246) = 1.14$, $p = 0.32$, the adjusted means of the variable from the model are presented in Table 6 and indicate that there is little difference between the means of each group. The centre variable is statistically significant at the 5% level $F(6, 246) = 2.50$, $p = 0.02$; post hoc tests would be required to evaluate between which centres the significant differences lie.

### QCPR at follow-up 2

On the QCPR at follow-up 2 the researcher variable is not statistically significant at the 5% level, $F(2, 234) = 2.93$, $p = 0.06$. The adjusted means of each level of the variable are presented in Table 6. The interaction between centre and researcher attendance is significant at the 5% level $F(10, 234) = 2.16$, $p = 0.02$, post hoc tests of this result are detailed in the interaction post hoc test results section below. The age of the participant is also significant on the QCPR follow-up 2 variable at the 5% level $F(1, 234) = 6.01$, $p = 0.02$.

### Post hoc tests and further investigation of significant findings

The researcher attendance variable is significant at follow-up 1 on both outcome measures. The mean differences and effect sizes obtained were very small indicating that the mean difference is trivial but there is an effect of the variable in the model (Cohen, 1988). As an interaction term between the covariate and independent variable is included in the model then the interpretation of the results should be treated in relation to this interaction (Grace-Martin, 2013). Presented in Figures 7 and 8 are scatter diagrams of the baseline outcome measure values and the follow-up 1 scores split by researcher attendance groups. The graphs indicate that overall, the baseline scores have a positive relationship with the follow-up scores with a $R^2$ coefficient of 0.31 on the QCPR and $R^2 = 0.43$ on the QoL-AD measure. When split by the researcher attendance groups the relationship between baseline

and follow-up scores is stronger in the same researcher group compared in the different researchers group. For the QCPR measure the same researcher group has a medium-large coefficient of $R^2 = 0.49$ compared to the different researcher groups having a small-medium r squared coefficient of $R^2 = 0.16$. Similarly, on the QoL-AD the same researcher group has a large coefficient of $R^2 = 0.59$ and the different researchers group has a small-medium coefficient of $R^2 = 0.29$ which is a small-medium effect size. This is the case for both the QoL-AD and QCPR but is more prominent for QCPR scores, with the QCPR having a bigger difference in $R^2$ coefficients between the two groups.

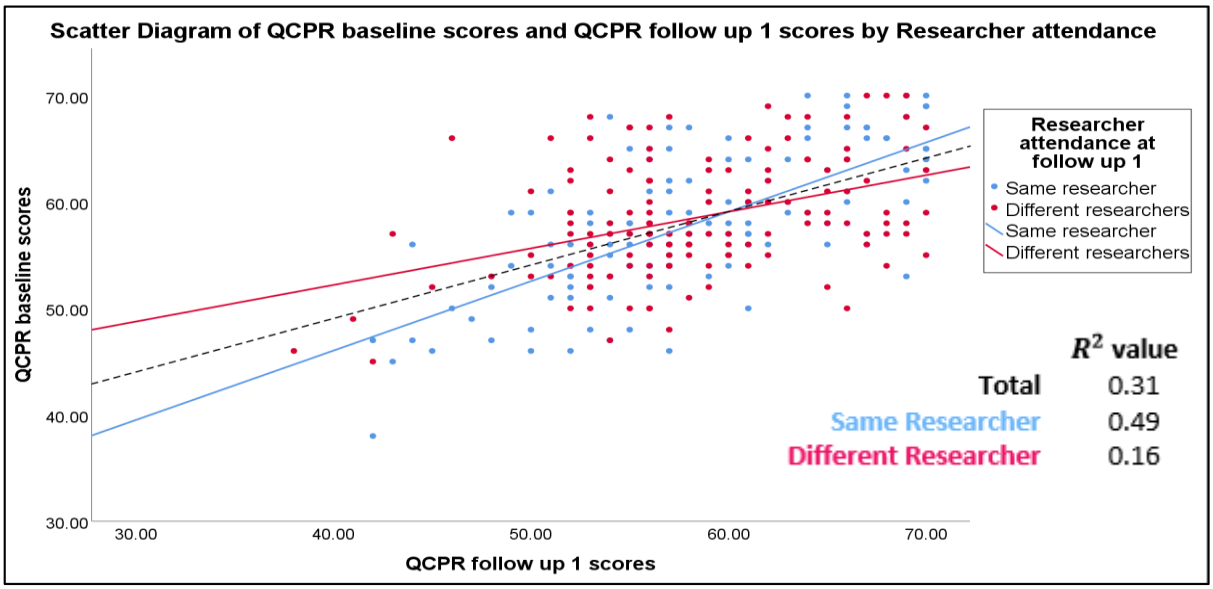


***Figure 8:*** *Scatter Plot of QCPR Baseline and Follow-up 1 Scores, split by Researcher Attendance*

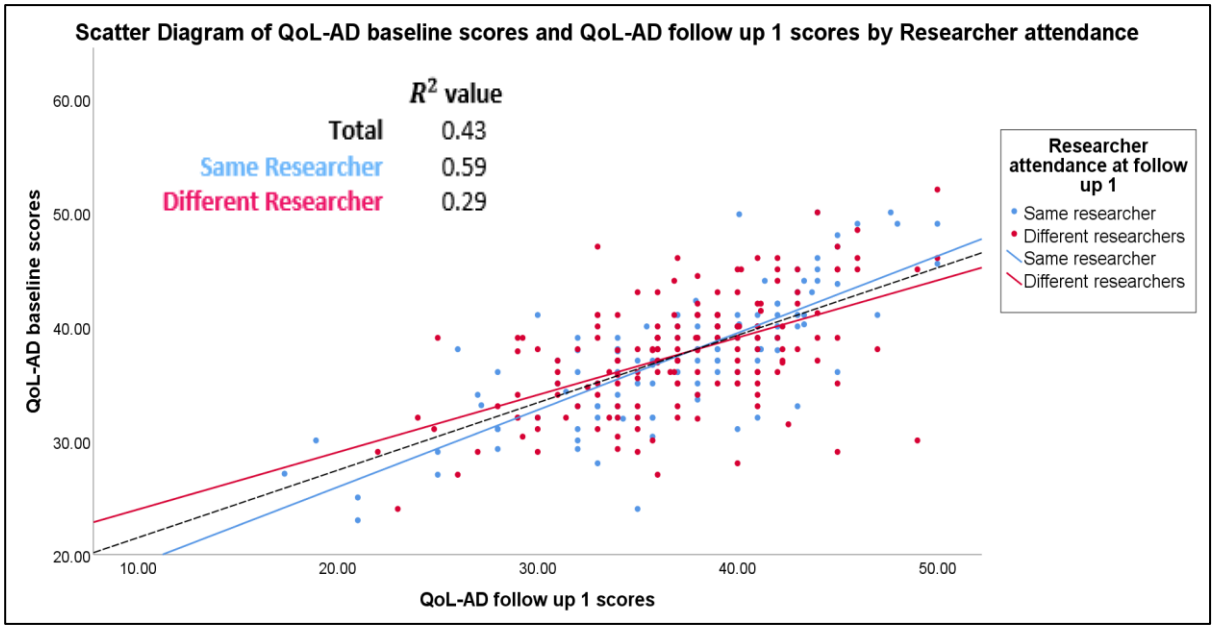


***Figure 7:*** *Scatter plot of QoLAD baseline and Follow-up 1 scores, Split by Researcher Attendance*

Interaction Post hoc tests

As the interaction between centre and researcher attendance was significant on the QCPR outcome measure at follow-up 2, post hoc pairwise comparison tests have been conducted and are presented in Table 7. Once split into researcher attendance group at each centre there are very small samples representing some groups, therefore the results are interpreted with caution.

The pairwise comparisons revealed that there is a statistically significant mean difference of 6.84, $p < 0.01$, at the 1% level at the Hull centre between the two different and three different researcher groups. A Cohen's *d* effect size of 1.10 was calculated for the difference indicating a large effect (Cohen, 1988), with those who had three different researchers having an overall higher score on the QCPR. At the London North Centre a statistically significant mean difference of 4.03 at the 5% level ($p = 0.05$) was found, with a Cohen's *d* medium effect size of -0.58 (Cohen, 1988). This difference again is between those who had two different and three different researchers - those who had two different had a higher overall mean score. Lastly, a mean difference of 5.84 was found to be statistically significant at the 5% level at the Manchester centre between those in the same researcher group and those in the three different researchers group. This difference produced a Cohen's *d* medium effect size of -0.56 with those who had the same researcher having a higher overall mean than those who had three different (Cohen, 1988).

The differences are illustrated graphically in Figure 9 with a scatter plot of the mean QCPR follow-up 2 scores at each centre split by the researcher attendance group along with the associated 95% Confidence interval error bars. The confidence interval ranges vary and are noticeably large at Gwent for the same researcher and three different researcher groups, at Manchester for the three different and at London North for the same researcher group. The large confidence intervals indicate that the estimated marginal means obtained may not be reliable, which further highlights the need to interpret the results with caution due to the small sample sizes representing some of the groups, see Table 7.

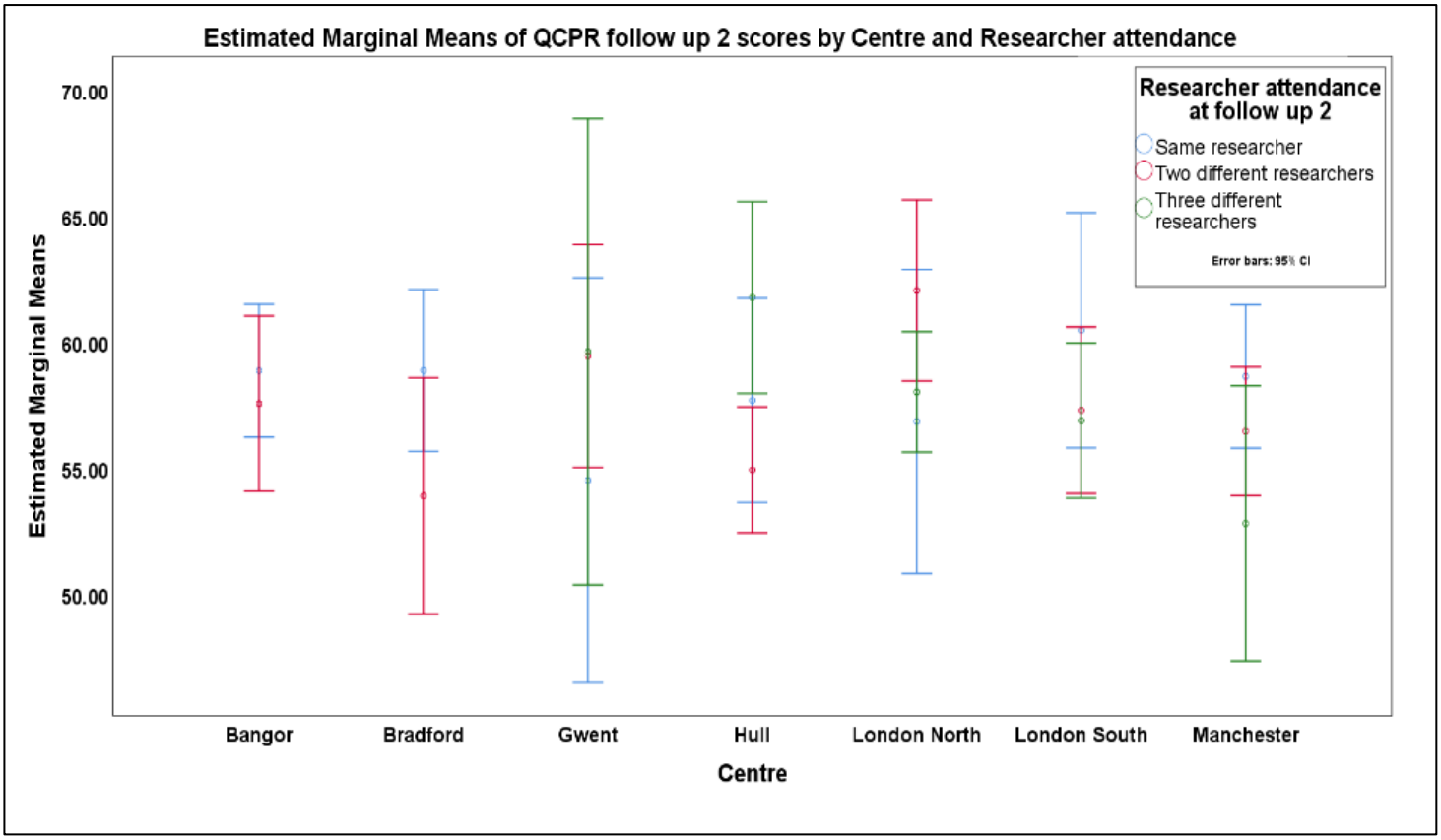


***Figure 9:*** *Plot of Estimated Marginal Means of QCPR Follow-up 2 at Each Site, by Researcher Attendance*

***Table 7:*** *QCPR Follow-up 2 Pairwise Comparison Tests of Centre and Researcher Attendance*

| CENTRE | RESEARCHER ATTENDANCE AT FU2 | ADJUSTED VALUES | | RESEARCHER ATTENDANCE AT FU2 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | *SAME RESEARCHER* | | | *TWO DIFFERENT* | | | *THREE DIFFERENT* | | |
| | | N | MEAN (SE) | MEAN DIFF | SIG | EFFECT SIZE (95% CI) | MEAN DIFF | SIG | EFFECT SIZE (95% CI) | MEAN DIFF | SIG | EFFECT SIZE (95% CI) |
| **BANGOR** | ***Same*** | 28 | 58.8 (1.34) | | | | 1.31 | p = 0.52 | -0.19 (-0.83, 0.46) | | | |
| | ***Two different*** | 14 | 57.5 (1.77) | -1.31 | p = 0.52 | 0.19 (-0.46, 0.83) | | | | N/A | | |
| | ***Three different*** | 0 | N/A | N/A | | | N/A | | | | | |
| **BRADFORD** | ***Same*** | 22 | 58.9 (1.63) | | | | 4.98 | p = 0.07 | -0.67 (-1.46, 0.13) | | | |
| | ***Two different*** | 9 | 53.9 (2.38) | -4.98 | p = 0.07 | 0.67 (-0.13, 1.46) | | | | N/A | | |
| | ***Three different*** | 0 | N/A | N/A | | | N/A | | | | | |
| **GWENT** | ***Same*** | 3 | 54.5 (4.08) | | | | -4.93 | p = 0.25 | 0.75 (-0.61, 2.12) | -5.10 | p = 0.34 | 0.67 (-0.98, 2.31) |
| | ***Two different*** | 8 | 59.4 (2.24) | 4.93 | p = 0.25 | -0.75 (-2.12, 0.61) | | | | -0.17 | p = 0.97 | 0.03 (-1.30, 1.36) |
| | ***Three different*** | 3 | 59.6 (4.70) | 5.10 | p = 0.34 | -0.67 (-2.31, 0.98) | 0.17 | p = 0.97 | -0.03 (-1.36, 1.30) | | | |
| **HULL** | ***Same*** | 11 | 57.7 (2.06) | | | | 2.76 | p = 0.25 | -0.44 (-1.17, 0.28) | -4.08 | p = 0.14 | 0.60 (-0.25, 1.46) |
| | ***Two different*** | 23 | 54.9 (1.27) | -2.76 | p = 0.25 | 0.44 (-0.28, 1.17) | | | | **-6.84 | p < 0.01 | 1.10 (0.33, 1.86) |
| | ***Three different*** | 11 | 61.7 (1.93) | 4.08 | p = 0.14 | -0.60 (-1.46, 0.25) | **6.84 | p < 0.01 | -1.10 (-1.86, -0.33) | | | |
| **LONDON NORTH** | ***Same*** | 4 | 56.8 (3.06) | | | | -5.21 | p = 0.14 | 0.83 (-0.34, 1.99) | -1.18 | p = 0.71 | 0.17 (-0.87, 1.21) |
| | ***Two different*** | 12 | 62.0 (1.83) | 5.21 | p = 0.14 | -0.83 (-1.99, 0.34) | | | | *4.03 | p = 0.05 | -0.58 (-1.25, 0.09) |
| | ***Three different*** | 34 | 58.0 (1.21) | 1.18 | p = 0.71 | -0.17 (-1.21, 0.87) | *-4.03 | p = 0.05 | 0.58 (-0.09, 1.25) | | | |
| **LONDON SOUTH** | ***Same*** | 7 | 60.4 (2.37) | | | | 3.17 | p = 0.26 | -0.47 (-1.37, 0.43) | 3.89 | p = 0.19 | -0.48 (-1.34, 0.37) |
| | ***Two different*** | 16 | 57.3 (1.68) | -3.17 | p = 0.26 | 0.47 (-0.43, 1.37) | | | | 0.41 | p = 0.85 | -0.06 (-0.69, 0.58) |
| | ***Three different*** | 23 | 56.9 (1.56) | -3.89 | p = 0.19 | 0.48 (-0.37, 1.34) | -0.41 | p = 0.85 | 0.06 (-0.58, 0.69) | | | |
| **MANCHESTER** | ***Same*** | 23 | 56.6 (1.44) | | | | 2.18 | p = 0.20 | -0.03 (-0.56, 0.51) | *5.84 | p = 0.05 | -0.56 (-1.54, 0.42) |
| | ***Two different*** | 32 | 56.4 (1.30) | -2.18 | p = 0.20 | 0.03 (-0.51, 0.56) | | | | 3.66 | p = 0.21 | -0.50 (-1.45, 0.45) |
| | ***Three different*** | 5 | 52.8 (2.77) | *-5.84 | p = 0.05 | 0.56 (-0.42, 1.54) | -3.66 | p = 0.21 | 0.50 (-0.45, 1.45) | | | |

*Significant at 5% level, ** Significant at 1% level

### 4.2.2 Carer data results

The results from the ANCOVA models run on the carer data are displayed in Table 8 and the adjusted means for the follow-up scores for each level of the researcher attendance variable are presented in Table 9.

#### QCPR follow-up 1

The researcher attendance variable is not statistically significant on the QCPR measure at follow-up 1 $F(1, 270) = 0.03$, $p = 0.86$, and neither is the interaction between centre and the researcher attendance variable $F(6, 270) = 0.81$, $p = 0.57$.

The carer's marital status $F(1, 270) = 5.18$, $p = 0.02$, and the group allocation $F(1, 270) = 3.89$, $p = 0.05$, are statistically significant at the 5% level. On average the marital status group 'spousal' have a higher mean score on the QCPR indicating a better perceived relationship between carer and participant and, on average, the treatment as usual group have a higher mean on the QCPR score indicating a better perceived carer and participant relationship.

#### QCPR follow-up 2

There are no statistically significant findings at the 5% level on the QCPR at follow-up 2 for the researcher attendance variable for the main effect or the interaction with centre or on any other factors in the model.

#### QoL-AD proxy follow-up 1

The researcher attendance variable main effect is not statistically significant on the carer proxy QoL-AD at follow-up 1 $F(1, 285) = 0.70$, $p = 0.40$, neither was the interaction between centre and the researcher attendance variable $F(6, 285) = 0.61$, $p = 0.72$. However, the interaction between centre and allocation is statistically significant at the 5% level $F(6, 285) = 2.38$, $p = 0.03$.

#### QoL-AD proxy follow-up 2

On the Qol-AD at follow-up 2 the researcher attendance variable is statistically significant at the 5% level $F(2, 282) = 4.15$, $p = 0.02$. To evaluate where the significant differences lie and the magnitude of these differences post hoc tests have been conducted, the results of which (displayed in Appendix 8) found no statistically significant differences between any two levels of the researcher attendance groups further investigations of this effect are detailed in the next section 'Carer Post hoc tests and investigation of significant findings'.

***Table 8:*** *ANCOVA Model Results for Carer Data*

| QCPR follow-up 1 | | | | QoL-AD follow-up 1 | | | |
|---|---|---|---|---|---|---|---|
| **Factor** | **DF** | **F-value** | **p-value** | **Factor** | **DF** | **F-value** | **p-value** |
| **QCPR carer Baseline** | 1 | 288.79 | **<0.01 | **QoL-AD proxy Baseline** | 1 | 357.58 | **<0.01 |
| **PwD Age** | 1 | 0.76 | 0.39 | **PwD Age** | 1 | 1.06 | 0.30 |
| **Carer Gender** | 1 | 0.57 | 0.45 | **Carer Gender** | 1 | 0.15 | 0.70 |
| **Carer Age** | 1 | 0.20 | 0.66 | **Carer Age** | 1 | 0.03 | 0.87 |
| **PwD Gender** | 1 | 3.16 | 0.08 | **PwD Gender** | 1 | 0.20 | 0.66 |
| **Carer Marital status** | 1 | 5.18 | *0.02 | **Carer Marital status** | 1 | 0.43 | 0.51 |
| **Centre** | 6 | 0.32 | 0.93 | **Centre** | 6 | 0.51 | 0.80 |
| **Wave** | 4 | 0.46 | 0.76 | **Wave** | 4 | 1.26 | 0.29 |
| **Allocation** | 1 | 3.89 | *0.05 | **Allocation** | 1 | 2.10 | 0.15 |
| **Centre x Allocation** | 6 | 1.31 | 0.25 | **Centre x Allocation** | 6 | 2.38 | *0.03 |
| **Researcher Attendance** | 1 | 0.03 | 0.86 | **Researcher Attendance** | 1 | 0.70 | 0.40 |
| **Centre x Researcher Attendance** | 6 | 0.81 | 0.57 | **Centre x Researcher Attendance** | 6 | 0.61 | 0.72 |
| **Error (SS within)** | 270 | | | **Error (SS within)** | 285 | | |
| **QCPR follow-up 2** | | | | **QoL-AD follow-up 2** | | | |
| **Factor** | **DF** | **F-value** | **p-value** | **Factor** | **DF** | **F-value** | **p-value** |
| **QCPR carer Baseline** | 1 | 188.6 | **<0.01 | **QoL-AD proxy Baseline** | 1 | 230.61 | **<0.01 |
| **PwD Age** | 1 | 2.22 | 0.14 | **PwD Age** | 1 | 0.29 | 0.59 |
| **Carer Gender** | 1 | 1.27 | 0.26 | **Carer Gender** | 1 | 2.88 | 0.09 |
| **Carer Age** | 1 | 2.71 | 0.10 | **Carer Age** | 1 | 1.14 | 0.29 |
| **PwD Gender** | 1 | 0.25 | 0.62 | **PwD Gender** | 1 | 0.86 | 0.36 |
| **Carer Marital status** | 1 | 0.08 | 0.78 | **Carer Marital status** | 1 | 0.45 | 0.51 |
| **Centre** | 6 | 0.54 | 0.78 | **Centre** | 6 | 1.20 | 0.31 |
| **Wave** | 4 | 0.84 | 0.50 | **Wave** | 4 | 1.12 | 0.35 |
| **Allocation** | 1 | 0.00 | 0.99 | **Allocation** | 1 | 0.23 | 0.63 |
| **Centre x allocation** | 6 | 1.11 | 0.36 | **Centre x Allocation** | 6 | 0.49 | 0.82 |
| **Researcher Attendance** | 2 | 0.46 | 0.63 | **Researcher Attendance** | 2 | 4.15 | *0.02 |
| **Centre x Researcher Attendance** | 10 | 0.98 | 0.46 | **Centre x Researcher Attendance** | 10 | 0.94 | 0.50 |
| **Error (SS within)** | 272 | | | **PwD age*Researcher Attendance** | 2 | 3.96 | *0.02 |
| | | | | **Error (SS within)** | 282 | | |

**Significant at the 0.05 level. **Significant at the 0.01 level.*

***Table 9:*** *Estimated Marginal Means of Researcher Attendance*

| **OUTOME MEASURE** | **ADJUSTED VALUES AT FOLLOW-UP** | | | | | |
|---|---|---|---|---|---|---|
| | **Same Researcher** | | **Two Different Researchers** | | **Three Different Researchers** | |
| | **N** | **Mean (SE)** | **N** | **Mean (SE)** | **N** | **Mean (SE)** |
| **QCPR FU1** | 145 | 52.7 (1.00) | 156 | 52.9 (0.98) | N/A | |
| **QOL-AD FU1** | 153 | 31.9 (0.67) | 163 | 31.3 (0.67) | N/A | |
| **QCPR FU2** | 113 | 53.1 (1.46) | 123 | 53.0 (1.13) | 72 | 54.4 (1.65) |
| **QOL-AD FU2** | 114 | 31.7 (0.90) | 127 | 30.9 (0.69) | 79 | 31.8 (1.01) |

*Carer Post hoc tests* and investigation of significant findings

As the QoL-AD analysis included the interaction between the age of the person with Dementia and follow-up 2 scores the interpretation of the results should be treated in relation to this (Grace-Martin, 2013). The scatter diagram in Figure 10 shows the relationship between the participant age and the carer proxy QoL-AD follow-up scores. The overall line of fit (black line) reveals that there is a very weak association with higher participant age and lower proxy QoL-AD scores, with an $R^2$ coefficient of 0.02. When evaluating these lines at sub group levels the relationship is still a small association. It is stronger for those in the same researcher group ($R^2$= 0.07) than those in the two or three different researcher groups ($R^2$ = 0.008 and $R^2$ = 0.001 respectively).

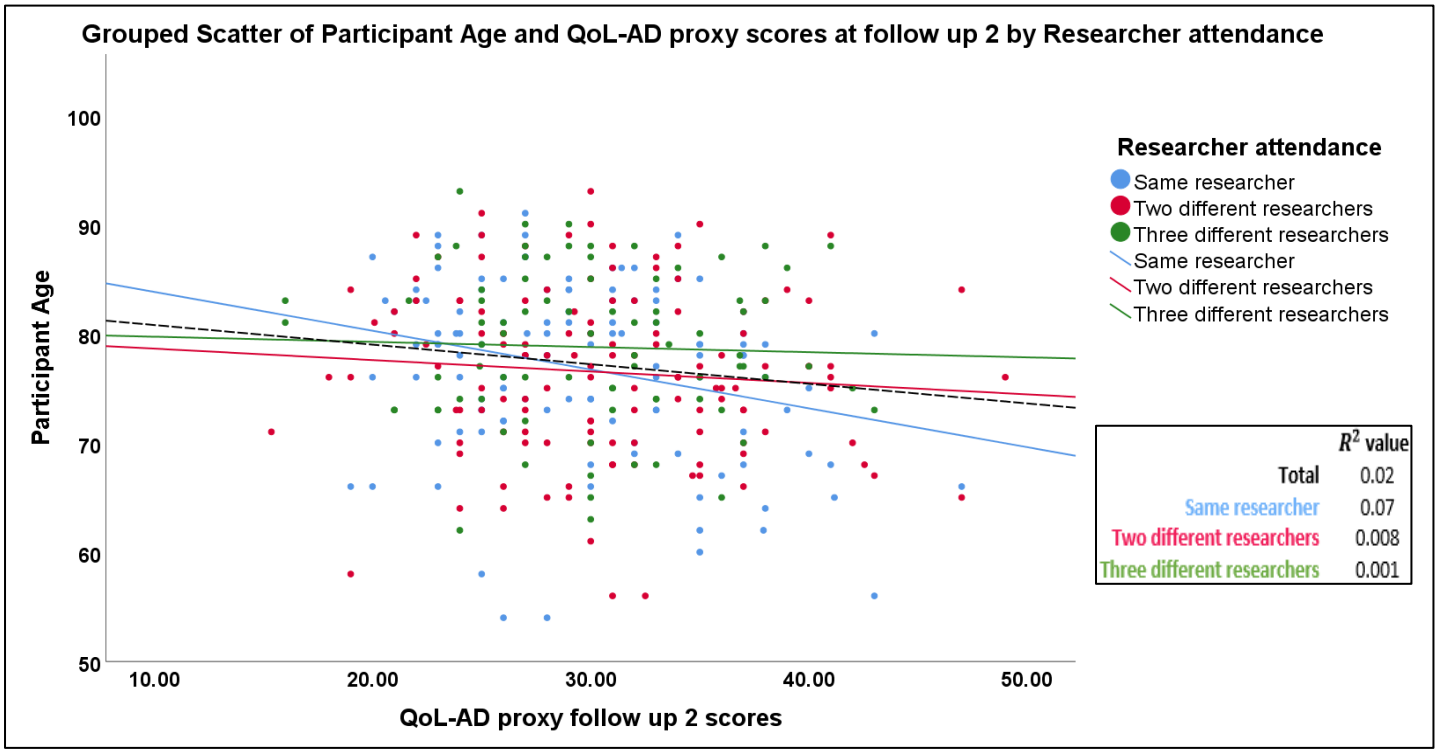


***Figure 10:*** *Scatter Plot of Carer QoL-AD Proxy Baseline and Follow-up 2, by Researcher Attendance*

## 4.3 Sensitivity Analysis

As described in Chapter Three, the researcher variable at follow-up 2 is split into three groups; same, two different and three different. Summarised in Figure 11 by representing the same researcher occurrence with an 'x' and different with a 'y' or 'z', those in the same researcher group have all three visits with the same researcher, those in the three different have all three visits with three different researchers and the participants in the two different researchers groups have two different. This two different researchers group can be formed in one of three ways; either the same researcher conducts

baseline and follow-up 1 and a different conducts follow-up 2 (X, X, Y), or the same researcher conducts follow-up 1 and 2 but a different researcher conducted baseline (Y, X, X). The third scenario is when the same researcher conducts baseline and follow-up 2 but a different one conducts follow-up 1 (X, Y, X).

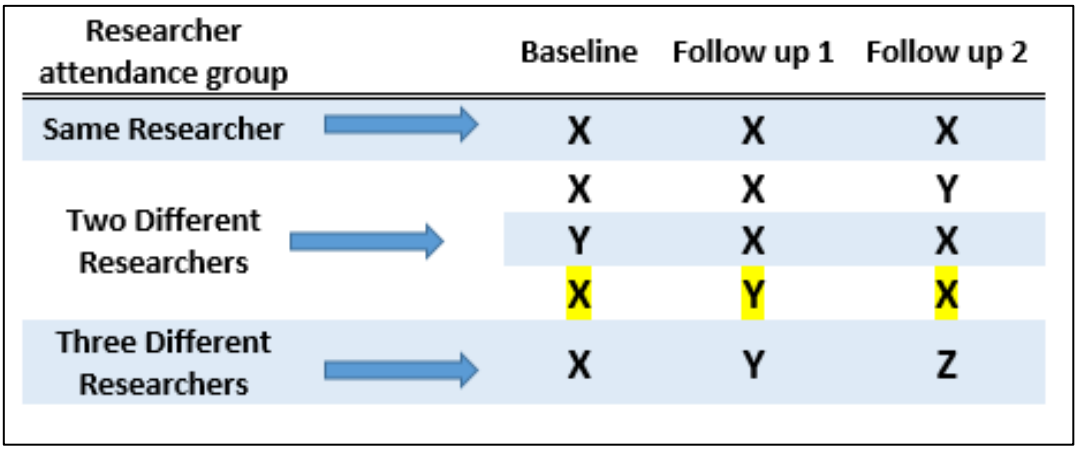


***Figure 11:*** *Grouping of researcher attendance variable based on researcher occurrence*

The idea behind the researcher influencing outcome measures scores is that there is a potential build-up of rapport between researcher and the PwD however in the scenario where there is a different researcher for the middle visit, this could potentially impact this relationship build-up. To investigate any possible impact of this coding on the analysis results a sensitivity analysis has been conducted re-coding those participants who had 'X, Y, X' pattern as highlighted in Figure 11 by placing the participants (PwD and their carer) that had this coding pattern into the 'three different researcher groups' since having this break of continuity could have a similar impact on a participant as they might perceive this as having three different researchers. Table 10 below summarises the re-coding figures for the researcher attendance groups.

***Table 10:*** *Recoding Figures of the Researcher Attendance Variable*

| | Same researcher N (previous N) | Two different N (previous N) | Three Different N (previous N) |
|---|---|---|---|
| **Overall** | 118 (118) | 113 (129) | 99 (83) |
| **QCPR participant** | 91 (91) | 99 (115) | 90 (74) |
| **QoL-AD participant** | 92 (92) | 94 (108) | 84 (70) |
| **Carer QCPR** | 113 (113) | 107 (123) | 88 (72) |
| **Carer QoL-AD proxy** | 114 (114) | 111 (127) | 95 (79) |

All the participants in the same researcher group were not re-coded, therefore, as expected, the numbers in the same researcher group overall and for each measure have not changed. Overall 16 cases were moved from the two to three group, for the outcome measures the participant QCPR, carer QCPR and carer QoL-AD proxy had all 16 cases complete these measures and included in analysis, however, for the participant QoL-AD two participants that were recoded had missing data therefore only 14 cases were affected by the recode in the final analysis models.

As with the primary analyses, the models were run as complete case and the assumptions of the models were re-checked with the recoded independent variable and resulting residuals of the models. The researcher attendance re-grouping had little effect on these assumptions, and all were evaluated to hold. As with the main analysis where the assumption of homogeneity of regression slopes was violated the interaction terms were included in the model (Grace-Martin, 2013). The only interaction affected by the regrouping was on the participant QoL-AD. For primary analysis there were no significant interactions between the covariates and independent variable. However, in the sensitivity model assumption checking the interaction between the recoded researcher attendance variable and the PwD age was significant and hence included in the analysis model.

### Sensitivity Analysis Results

The results of the sensitivity analysis are contained in Table 11. The aim of the sensitivity analysis is to explore whether the recoding of participants affects the results of the researcher attendance variable. The adjusted means of the measures presented for each researcher attendance recoded group are displayed in Table 12.

***Table 11:*** *ANCOVA Model Sensitivity Analysis Results at Follow-up 2*

| Participant QCPR follow-up 2 | | | | Participant QoL-AD follow-up 2 | | | |
|---|---|---|---|---|---|---|---|
| **Factor** | **DF** | **F-value** | **SIG** | **Factor** | **DF** | **F-value** | **SIG** |
| **QCPR Baseline** | 1 | 47.66 | **<0.01 | **QoL-AD Baseline** | 1 | 91.01 | **<0.01 |
| **Age** | 1 | 5.32 | *0.02 | **Age** | 1 | 0.75 | 0.39 |
| **Gender** | 1 | 0.02 | 0.89 | **Gender** | 1 | 0.22 | 0.64 |
| **Marital status** | 1 | 0.42 | 0.52 | **Marital status** | 1 | 0.56 | 0.46 |
| **Centre** | 6 | 0.84 | 0.54 | **Centre** | 6 | 3.06 | **0.01 |
| **Wave** | 4 | 0.42 | 0.80 | **Wave** | 4 | 1.08 | 0.37 |
| **Allocation** | 1 | 0.56 | 0.45 | **Allocation** | 1 | 0.33 | 0.57 |
| **Centre x Allocation** | 6 | 0.78 | 0.59 | **Centre x Allocation** | 6 | 0.91 | 0.49 |
| **Researcher Attendance** | 2 | 4.02 | *0.02 | **Researcher Attendance** | 2 | 3.32 | *0.04 |
| **Centre x Researcher Attendance** | 11 | 1.53 | 0.12 | **Centre x Researcher Attendance (fu2)** | 11 | 0.89 | 0.55 |
| **Baseline x Researcher Attendance** | 2 | 4.35 | **0.01 | **PwD Age x Researcher Attendance** | 2 | 2.96 | *0.05 |
| **Error (SS within)** | 233 | | | **Error (SS within)** | 280 | | |
| **Carer QCPR follow-up 2** | | | | **Carer QoL-AD proxy follow-up 2** | | | |
| **Factor** | **DF** | **F-value** | **SIG** | **Factor** | **DF** | **F-value** | **SIG** |
| **QCPR carer Baseline** | 1 | 184.46 | **<0.01 | **QoL-AD proxy Baseline** | 1 | 232.93 | **<0.01 |
| **PwD Age** | 1 | 1.74 | 0.19 | **PwD Age** | 1 | 0.28 | 0.60 |
| **Carer Gender** | 1 | 0.99 | 0.32 | **Carer Gender** | 1 | 3.22 | 0.07 |
| **Carer Age** | 1 | 2.87 | 0.09 | **Carer Age** | 1 | 1.26 | 0.26 |
| **PwD Gender** | 1 | 0.31 | 0.58 | **PwD Gender** | 1 | 1.27 | 0.26 |
| **Carer Marital status** | 1 | 0.06 | 0.81 | **Carer Marital status** | 1 | 0.71 | 0.40 |
| **Centre** | 6 | 0.77 | 0.59 | **Centre** | 6 | 1.29 | 0.26 |
| **Wave** | 4 | 1.34 | 0.26 | **Wave** | 4 | 0.96 | 0.43 |
| **Allocation** | 1 | 0.02 | 0.89 | **Allocation** | 1 | 0.33 | 0.57 |
| **Centre x Allocation** | 6 | 1.08 | 0.37 | **Centre x Allocation** | 6 | 0.55 | 0.77 |
| **Researcher Attendance** | 2 | 1.26 | 0.29 | **Researcher Attendance** | 2 | 4.05 | *0.02 |
| **Centre x Researcher Attendance** | 11 | 0.76 | 0.68 | **Centre x Researcher Attendance** | 11 | 1.12 | 0.35 |
| **Error (SS within)** | 271 | | | **PwD age x Researcher Attendance** | 2 | 3.98 | *0.02 |
| | | | | **Error (SS within)** | 281 | | |

**Significant at the 0.05 level. **Significant at the 0.01 level.*

***Table 12:*** *Estimated Marginal Means of Researcher Attendance*

| OUTOME MEASURE | ADJUSTED VALUES AT FOLLOW-UP | | | | | |
|---|---|---|---|---|---|---|
| | Same Researcher | | Two Different Researchers | | Three Different Researchers | |
| | N | Mean (SE) | N | Mean (SE) | N | Mean (SE) |
| **PWD QCPR FU2** | 92 | 57.9 (1.11) | 94 | 57.4 (0.90) | 84 | 58.3 (1.24) |
| **PWD QOL-AD FU2** | 91 | 36.8 (0.87) | 99 | 38.1 (0.70) | 90 | 37.5 (0.88) |
| **CARER QCPR FU2** | 113 | 53.1 (1.47) | 107 | 52.4 (1.15) | 88 | 54.5 (1.51) |
| **CARER PROXY QOL-AD FU2** | 114 | 31.8 (0.90) | 111 | 30.7 (0.70) | 95 | 32.0 (0.92) |

### Participant QoL-AD at follow-up 2

The researcher attendance variable on the participant QoL-AD was not significant when analysed as primarily coded but following recoding of the variable for sensitivity analysis it is significant at the 5% level on the participant QoL-AD at follow-up 2 F(2, 280) = 3.32, (p=0.04). The interaction between participant age and the researcher attendance variable, as indicated in the assumption checking, is also now significant at the 5% level but was not in the primary analysis. This difference is further investigated in the 'Sensitivity Analysis Post hoc tests' section below.

### Participant QCPR at follow-up 2

The researcher attendance variable was not statistically significant at follow-up 2 in the primary model but is significant at the 5% level $F(2, 233) = 4.02$, $p = 0.02$, for the recoded sensitivity analysis. The interaction between baseline and the researcher variable at follow-up 2 for the main analysis was significant at the 5% level $F(2, 234) = 3.15$, $p=0.05$, but in the sensitivity analysis is significant at the 1% level $F(2, 233) = 4.35$, $p = 0.01$. Another discrepancy between the two analyses is that the interaction between centre and researcher attendance is significant for the main analysis $F(10, 234) = 2.16$, $p=0.02$, but following recoding is no longer statistically significant in the sensitivity model $F(11, 233) = 1.53$, $p=0.12$.

### Carer QoL-AD and QCPR at follow-up 2

The results from the sensitivity analysis on the two carer outcomes at follow-up 2 produced no different findings on the researcher attendance variable to the primary models, suggesting that the recoding of the researcher attendance variable had no effect on the carer data. No factors are significant on the QCPR follow-up 2 for both the primary analysis and sensitivity analysis. On the QoL-AD proxy measure the researcher attendance along with the interaction between participant age and the researcher attendance variable are significant at the 2% level for both the main and sensitivity analyses.

### Sensitivity Analysis Post hoc tests

The researcher attendance variable is significant in the sensitivity analyses models on all the outcome measures, except for the carer QCPR, and as these include more than two groups (three groups) post hoc tests are required to investigate these effects further. Results of pairwise comparison tests are detailed in Appendix 9. The pairwise comparison tests' individual mean differences are not significant between any two groups. However, for all of the outcomes where the researcher attendance variable is significant, an interaction between the researcher attendance group and a

covariate was significant and included in the model, therefore the interpretation of the results consider these interactions (Grace-Martin, 2013).

Figure 12 displays the scatter diagram of the participant QCPR follow-up 2 scores against the QCPR baseline scores split by researcher attendance groups. The diagram indicates that there is a positive relationship between the QCPR baseline scores and the QCPR follow-up scores $R^2 = 0.13$ which is regarded as a small association (Cohen, 1988). This relationship is again strongest in the same researcher group which has a medium $R^2$ coefficient ($R^2 = 0.31$) compared with the two different researchers group which has a very weak relationship ($R^2 = 0.02$) and three different researcher groups with a small-medium relationship ($R^2 = 0.19$).

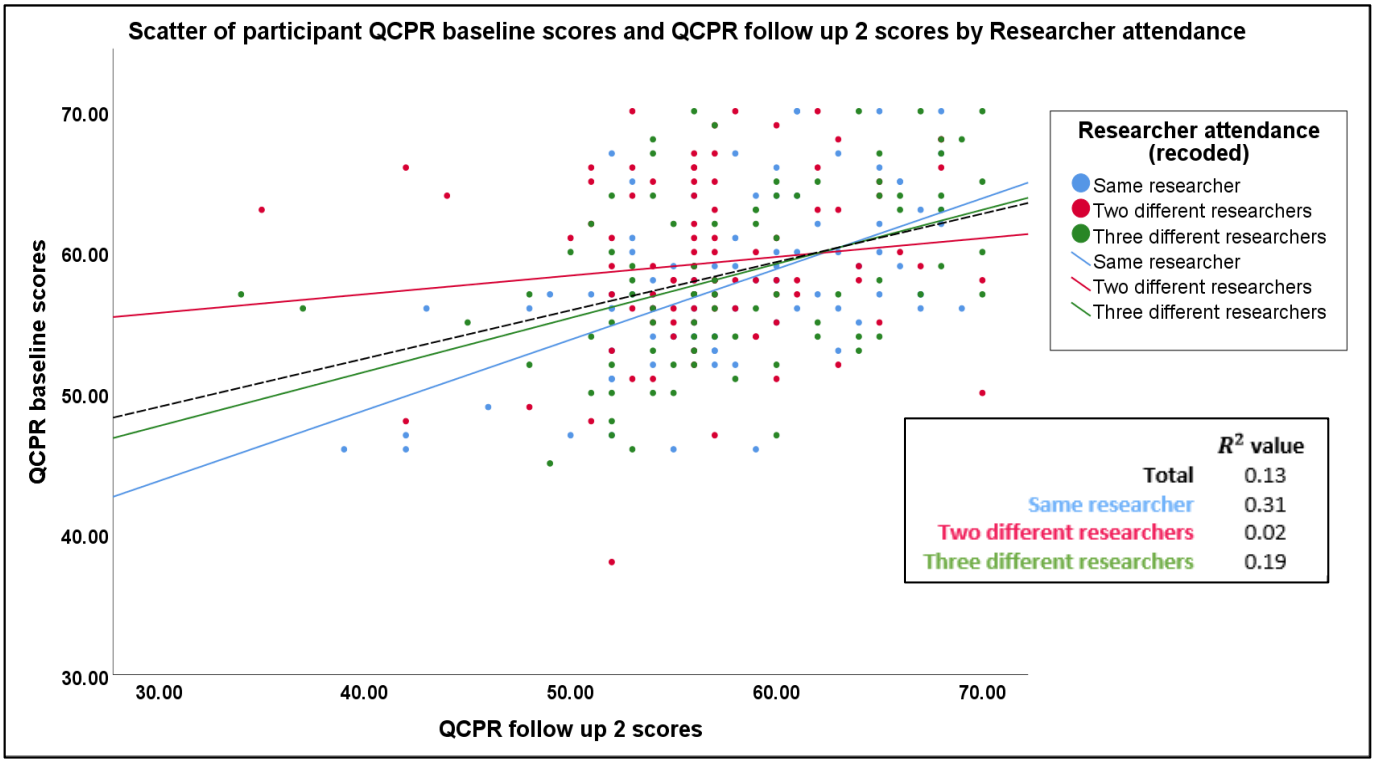


***Figure 12:*** *Scatter Plot of PwD QCPR Baseline and Follow-up 2, by Researcher Attendance (Recoded)*

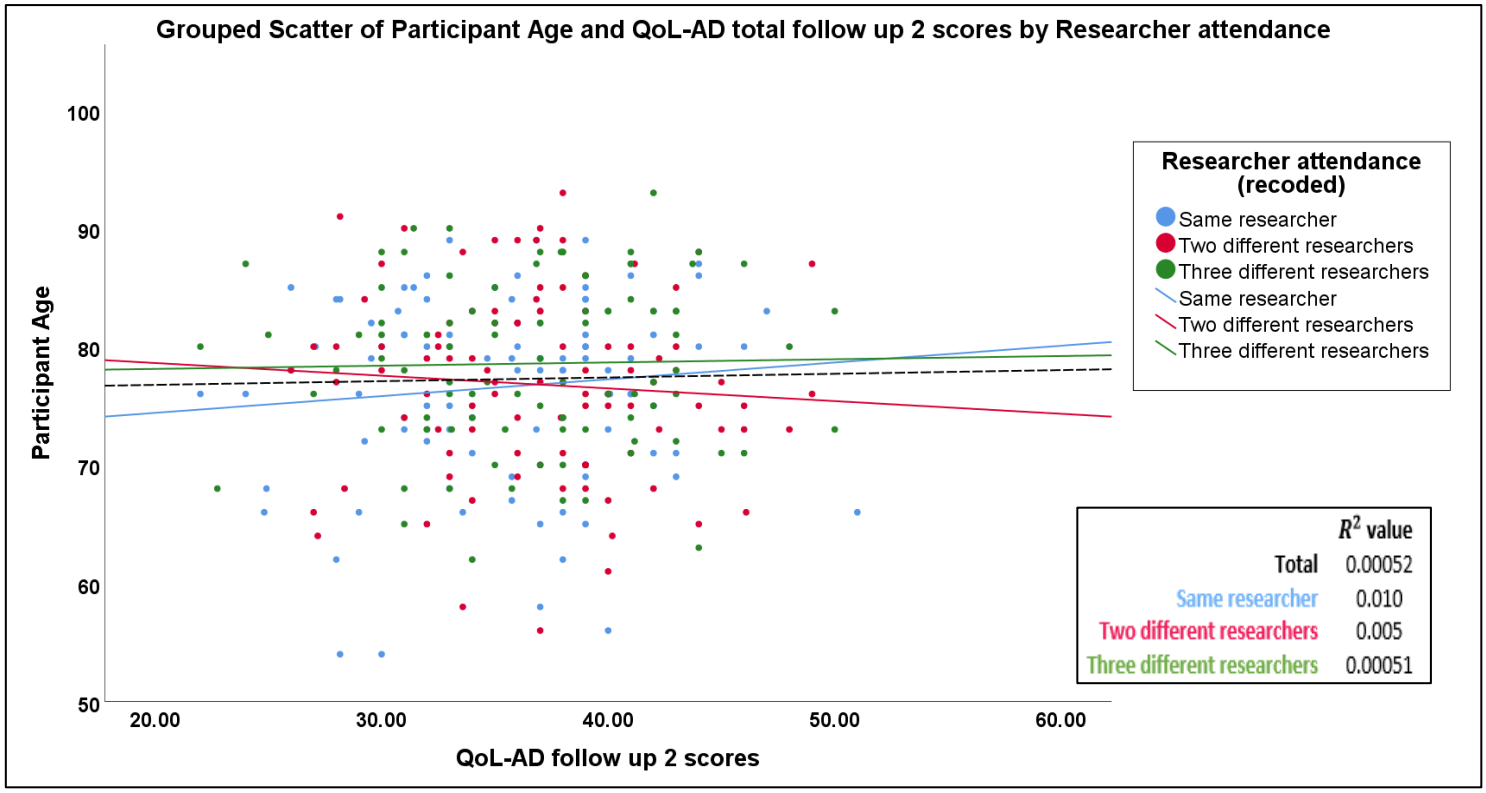


***Figure 13:*** *Scatter Plot of QoL-AD Follow-up 2 and PwD Age, by Researcher Attendance (recoded)*

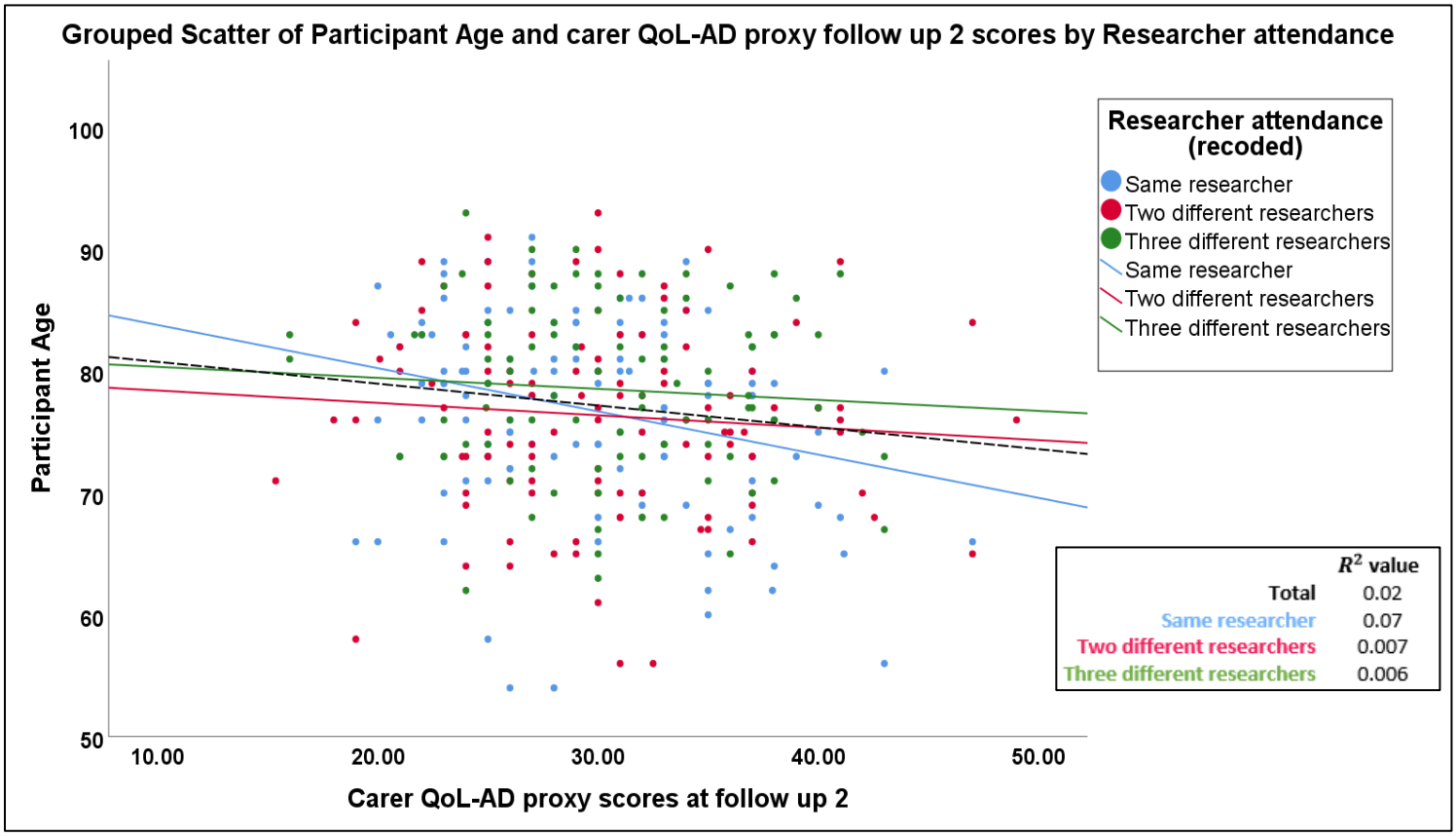


***Figure 14:*** *Scatter Plot of Carer PwD QoL-AD Follow-up 2 and PwD Age, by Researcher Attendance (recoded)*

The participant QoL-AD and carer proxy QoL-AD both had the interaction between PwD Age and researcher attendance included in the model, Figure 13 shows the participant QoL-AD scores at follow-up 2 against the participant age split by researcher group. The overall line of fit for the participant data shows a very weak relationship between the participant age and follow-up scores ($R^2 < 0.0001$). When split by researcher attendance groups those who had same researcher have a slightly stronger relationship between the two factors ($R^2 = 0.01$) however this is still a very weak association. The two different researcher group again has a weak $R^2$ coefficient ($R^2 = 0.005$) however it appears to be negative between factors whereas the same researcher group has a positive relationship between them. The three different researchers group has a similar relationship to the overall fit line with a very weak relationship ($R^2 < 0.0002$).

Figure 14 displays the carer QoL-AD proxy scores split by researcher attendance groups. The overall line of fit reveals that higher participant age is associated with a lower proxy QoL-AD rating ($R^2 = 0.02$), but when evaluating these relationships by each researcher attendance group this association is stronger for those in the same researcher group ($R^2 = 0.07$) than those in the two ($R^2 = 0.007$) or three different researcher group ($R^2 = 0.006$). The relationships overall and at sub groups are all very weak associations (Cohen, 1988). Interestingly, as with the participant QoL-AD scores, overall the same researcher group has a negative association between participant age and follow-up 2 scores whereas the two different researcher and three different researcher groups have a weaker negative relationship.

### Sensitivity analysis summary

The results obtained from the sensitivity analysis indicate that there are discrepancies between the results based on how the researcher attendance variable is coded at follow-up 2. Discrepancies were noted between primary and sensitivity models on both participant outcome measures on the researcher attendance main effect, with the variable not being significant in analysis with the primary coding but is statistically significant in the re-coding sensitivity analysis. There were also several discrepancies identified amongst the interaction terms of the researcher attendance variable and other factors (e.g. Centre, baseline and PwD Age). The results of the analyses suggest that those participants who had the same researcher then a different and then the same as baseline (x, y, x pattern in Figure 11) behave similarly to those who have three different (x, y, z).

The carer analysis results were not impacted by the recoding of the researcher attendance variable with no discrepancies noted to the statistical significance of the researcher attendance

variable in the sensitivity analysis compared with the primary results. This lack of distinction between the two carer analyses is logical in relation to the study hypothesis of an expected researcher effect. As the carers completed the assessments independently of the researcher and had little interaction with researchers when compared to the PwD, we would expect to see an effect in the PwD results and not the carer - therefore, the re-coding having no effect on the carer analysis aligns with this argument. The PwD sensitivity analysis indicates that there is some impact on the PwD measures indicating that the relationship (or something being represented by the variable) is affecting the measurement.

## 4.4 Chapter Summary

The primary analyses revealed that at follow-up 1, the researcher attendance variable was statistically significant at the 1% level of significance on the QoL-AD outcome and at the 5% level on the QCPR outcome. At follow-up 2, the researcher attendance variable was not significant on either outcome measures at the 1% or 5% level of significance. Sensitivity analysis on the coding of the researcher attendance factor at follow-up 2 showed that when the factor was recoded slightly differently the results of the participant analysis changed with the variable now being statistically significant in both the QoL-AD and QCPR follow-up 2 models at the 5% level of significance. This indicates that the way in which the variable is coded at follow-up may have an impact on findings within the participant data. Table 13 below summarises the main findings of the current Chapter.

***Table 13**: Summary of results of analysis for researcher in attendance variable*

| | Outcome | Main analysis | Sensitivity analysis |
|---|---|---|---|
| **Participant** | **QCPR follow up 1** | Significant at 5% level (p=0.02) | n/a |
| | **QoL-AD follow up 1** | Significant at 1% level (p<0.01) | n/a |
| | **QCPR follow up 2** | Not Significant (p=0.06) | Significant at 5% level (p=0.02) |
| | **QoL-AD follow up 2** | Not Significant (p=0.32) | Significant at 5% level (p=0.04) |
| **Carer** | **QCPR follow up 1** | Not Significant (p=0.86) | n/a |
| | **QoL-AD follow up 1** | Not Significant (p=0.40) | n/a |
| | **QCPR follow up 2** | Not Significant (p=0.63) | Not Significant (p=0.29) |
| | **QoL-AD follow up 2** | Significant at 5% level (p=0.02) | Significant at 5% level (p=0.02) |

The primary carer analyses indicated that the researcher attendance variable is not statistically significant on the QoL-AD at follow-up 1 nor on the QCPR at follow-up 1 or 2. There is, however, a statistically significant result on the QoL-AD proxy at follow-up 2 at the 5% level. The

sensitivity analysis at follow-up 2 revealed the same results on the carer data indicating that the coding of the researcher attendance variable at follow-up 2 has little effect on the carer measures.

For all significant analysis, where an interaction term between the independent variable and covariates were included in the models, the interpretation of the findings were evaluated with respect to these interactions. For the participant data this included the interactions between baseline and researcher attendance variables on the primary analyses at follow-up 1 on the QCPR and QoL-AD and on the sensitivity analyses at follow-up 2 on the QCPR. For all these findings, the post hoc evaluations revealed that the relationship between baseline scores and follow-up scores are stronger in the group where the participant had the same researcher compared with the different researcher groups (two or three). The PwD age and researcher attendance interactions were included in the participant sensitivity analysis on the QoL-AD at follow-up 2 and on the carer primary and sensitivity analysis at follow-up 2 on the QoL-AD proxy measure. Again, the relationship between the PwD age variable and the outcome follow-up scores is *slightly* stronger for those in the same researcher attendance group than those who had different for both the participant QoL-AD and the carer proxy QoL-AD.

# Chapter Five – Analysis of Gender of Researcher in Attendance (Research Question 2)

This chapter addresses the second research question of 'In a study with three time points; does the gender of the attending researcher impact upon the outcome measure?' To evaluate this, the aim of the analysis is to explore whether there are any statistical differences in follow-up scores for the participants who had the same researchers all of the same gender, different researchers but of the same gender or a combination of different researchers and genders. Analysing whether there is a difference between the groups will give an indication whether or not there is potential presence of an effect of the researcher gender.

Chapter Three described the coding of the gender of researcher in attendance variable which is the attendance of the researcher conducting visits and whether the researchers were of different genders or the same (for those who had different researchers). Further descriptive data of the genders of researchers are detailed in Table 14, including the number of researchers that were male or female overall and at each time point and the number of visits conducted by male researchers or female. Further details of the 43 researchers (gender and number of visits each conducted split across time points) can be found in Appendix 10, Tables 1 and 2. As mentioned in Chapter Two, gender was categorised as the 'biological' gender of the researcher and/or participant and was regarded as a binary: male or female.

***Table 14:*** *Gender of Researchers Collecting Data and Visits Conducted*

| **Gender of researchers across the study** | **Overall N (%)** | **Baseline N (%)** | **Follow-up 1 N (%)** | **Follow-up 2 N (%)** |
|---|---|---|---|---|
| | **(N = 43)** | (N = 36) | (N = 34) | (N = 38) |
| ***Male*** | **6** | 5 | 6 | 5 |
| ***Female*** | **34** | 28 | 27 | 21 |
| ***Not Known*** | **3** | 3 | 1 | 0 |
| **Number of visits conducted by each gender** | **(N = 990)** | (N = 330) | (N = 330) | (N = 330) |
| ***Male*** | **199 (20%)** | 45 (14%) | 64 (19%) | 77 (23%) |
| ***Female*** | **787 (79%)** | 282 (86%) | 265 (80%) | 253 (77%) |
| ***Not Known*** | **4 (<1%)** | 3 (<1%) | 1 (<1%) | 0 (0) |

In total there were 43 researchers who conducted visits across the study, a high proportion of these researchers were female (79%) compared to males (14%). For 3 researchers their gender is unknown due to the gender not being identifiable by their first name (either there was only an initial detailed or the name given could not be categorised to a specific gender). In total 990 visits were carried out in the current data, 330 at each time point. As expected, due to the higher numbers of female researchers, the majority of visits across the study were conducted by a female researcher (79%).

## 5.1 Analysis methods

The same analysis methods as described in Chapter Four (Section 4.1) have been adopted but with the gender of the researcher in attendance variable being explored. To evaluate any differences of the means of the outcome measures between the levels of the gender of researcher in attendance variable ANCOVA models were adopted. Again, eight separate models were run as listed in Figure 6 Section 4.1, which were built to correspond with the models run in the REMCARE study. All factors for the participant and carer data are the same as in the analysis of the first research question and all the models were run as complete case analyses without using any missing data imputation techniques. The assumptions associated with an ANCOVA model have been checked for all the models in the same way as with research question 1 using the gender of the researcher attendance as the independent variable. A full list of the assumptions and methods to check can be found in Appendix 6.

There are three unknown researcher genders in the dataset, of these, two conducted one interview each and the other missing gender researcher conducted two, therefore there are four observations that have unknown “gender of researcher in attendance” values. Therefore, the main analysis datasets contain 326 participants in each (PwD and carer). A sensitivity analysis has been carried out to check what impacts these unknown genders have on the results by including all unknowns firstly as males and conducting analysis and then assuming them to be female and running the models to compare the results. The results of these sensitivity analyses are detailed in Section 5.3.

## 5.2 Analysis Results

### Assumption checks

A linear relationship between each of the covariates and dependent variable on each level of the researcher attendance group was revealed by visual inspection of scatterplots for all models. Again, the assumption of homogeneity of regression slopes was violated in several cases. Where this

assumption was violated, as recommended by Grace-Martin (2013), the interaction term was included in the final model. The list of the interaction terms in the models are contained in Appendix 7 and indicate which models satisfied the assumption and which did not. The interaction terms that were significant and hence included in the main models are reported within the analysis results table.

There were no substantial outliers to consider. Any potential outliers observed in the data were all within expected range of the measures and should not be removed from the analysis. Neither are the deviations substantial enough to need to consider transformations of the data. The Q-Q plots of the residuals indicated that a "perfect" normal distribution was not present, however, the data is only slightly skewed at the tails, therefore, transformations to the data are not required and the assumption *sufficiently* holds. Lastly, the assumptions of homoscedasticity and homogeneity of variances were met for all models.

### 5.2.1 Participant (PwD) data results

Table 15 contains the results of the participant ANCOVA models run. The estimated marginal means, (having adjusted for the covariates in the models), are presented for the gender of researcher in attendance variable at follow-up 1 and 2 in Table 16. Where significance was indicated then the associated effect sizes and confidence intervals are presented in the table.

#### QOL-AD at follow-up 1

There is a statistically significant difference on the gender of researcher in attendance variable at the 1% level $F(2, 258) = 5.74$, $p = 0.01$, on the QoL-AD at follow-up 1. The pairwise comparisons conducted to assess the magnitude of the differences between the levels produced no statistically significant differences. Cohen's *d* effect sizes of 0.07 and 0.15 which are classed as a small effect sizes indicates that the difference, although is statistically significant, is minor and potentially clinically insignificant (Cohen, 1988; Walker, 2008).

***Table 15:*** *ANCOVA Model Results for QCPR and QoL-AD PwD Data*

| Factor | DF | F-value | Significance (p-value) | Factor | DF | F-value | Significance (p-value) |
|---|---|---|---|---|---|---|---|
| **QCPR follow-up 1** | | | | **QoL-AD follow-up 1** | | | |
| **QCPR Baseline** | 1 | 124.62 | **<0.01 | **QoL-AD Baseline** | 1 | 94.97 | **<0.01 |
| **Age** | 1 | 0.14 | 0.71 | **Age** | 1 | 0.83 | 0.36 |
| **Gender** | 1 | 0.16 | 0.69 | **Gender** | 1 | 2.18 | 0.14 |
| **Marital status** | 1 | 0.07 | 0.79 | **Marital status** | 1 | 0.11 | 0.75 |
| **Centre** | 6 | 0.30 | 0.94 | **Centre** | 6 | 1.70 | 0.12 |
| **Wave** | 4 | 1.76 | 0.14 | **Wave** | 4 | 0.55 | 0.70 |
| **Allocation** | 1 | 0.31 | 0.58 | **Allocation** | 1 | 0.63 | 0.43 |
| **Centre x Allocation** | 6 | 0.22 | 0.97 | **Centre x Allocation** | 6 | 2.08 | 0.06 |
| **Researcher Gender Attendance** | 2 | 3.29 | *0.04 | **Researcher Gender Attendance** | 2 | 5.47 | **0.01 |
| **Centre x Researcher Gender Attendance** | 8 | 0.59 | 0.78 | **Centre x Researcher Gender Attendance** | 9 | 0.74 | 0.68 |
| **Error (SS within)** | 240 | | | **QoL-AD baseline x Researcher Gender Attendance** | 2 | 4.78 | **0.01 |
| | | | | **Error (SS within)** | 258 | | |
| **QCPR follow-up 2** | | | | **QoL-AD follow-up 2** | | | |
| **QCPR Baseline** | 1 | 47.44 | **<0.01 | **QoL-AD Baseline** | 1 | 93.27 | **<0.01 |
| **Age** | 1 | 3.87 | *0.05 | **Age** | 1 | 0.35 | 0.55 |
| **Gender** | 1 | 0.004 | 0.95 | **Gender** | 1 | 0.14 | 0.70 |
| **Marital status** | 1 | 0.18 | 0.67 | **Marital status** | 1 | 0.68 | 0.41 |
| **Centre** | 6 | 0.47 | 0.83 | **Centre** | 6 | 2.26 | *0.04 |
| **Wave** | 4 | 0.39 | 0.82 | **Wave** | 4 | 0.58 | 0.68 |
| **Allocation** | 1 | 0.30 | 0.58 | **Allocation** | 1 | 0.17 | 0.68 |
| **Centre x Allocation** | 6 | 0.77 | 0.59 | **Centre x Allocation** | 6 | 0.69 | 0.66 |
| **Researcher Gender Attendance** | 2 | 3.90 | *0.02 | **Researcher Gender Attendance** | 2 | 1.37 | 0.26 |
| **Centre x Researcher Gender Attendance** | 9 | 1.67 | 0.09 | **Centre x Gender Researcher Attendance** | 9 | 0.86 | 0.56 |
| **QCPR baseline x Researcher Gender Attendance** | 2 | 3.65 | 0.03 | **Error (SS within)** | 244 | | |
| **Error (SS within)** | 232 | | | | | | |

**Significant at the 0.05 level. **Significant at the 0.01 level.*

***Table 16:*** *Estimated Marginal Means for Gender of Researcher in Attendance Variable*

| OUTCOME MEASURE AND VARIABLE LEVEL | ADJUSTED VALUES | | MEAN DIFFERENCES (P-VALUE OF DIFFERENCE) COHEN'S D EFFECT SIZE (95% CI) | | |
|---|---|---|---|---|---|
| | N | Mean (SE) | *Same Researcher Same gender* | *Different Researchers Same gender* | *Different Researchers Different Genders* |
| **QCPR follow-up 1** | | | | | |
| **Same Researcher Same Gender** | 132 | 58.2 (0.95) | | -0.83 (p = 0.48)<br>0.07 (-0.19, 0.33) | *2.60 (p = 0.03)<br>-0.26 (-0.61, 0.09) |
| **Different Researchers Same Gender** | 98 | 59.0 (1.14) | 0.83 (p = 0.48)<br>-0.07 (-0.33, 0.19) | | **3.44 (p = 0.01)<br>-0.34 (-0.70, 0.03) |
| **Different Researchers Different Genders** | 42 | 55.6 (1.02) | *-2.60 (p = 0.03)<br>0.26 (-0.09, 0.61) | **-3.44 (p = 0.01)<br>0.34 (-0.03, 0.70) | |
| **QOL-AD follow-up 1** | | | | | |
| **Same Researcher Same Gender** | 136 | 37.1 (0.66) | | -0.64 (p = 0.45)<br>0.07 (-0.18, 0.32) | -1.26 (p = 0.34)<br>0.15 (-0.19, 0.49) |
| **Different Researchers Same Gender** | 112 | 37.7 (0.83) | 0.64 (p = 0.45)<br>-0.07 (-0.32, 0.18) | | -0.63 (p = 0.66)<br>0.07 (-0.28, 0.42) |
| **Different Researchers Different Genders** | 45 | 38.3 (1.27) | 1.26 (p = 0.34)<br>-0.15 (-0.49, 0.19) | 0.63 (p = 0.66)<br>-0.07 (-0.42, 0.28) | |
| **QCPR follow-up 2** | | | | | |
| **Same Researcher Same Gender** | 92 | 58.0 (1.11) | | 0.37 (p = 0.76)<br>-0.03 (-0.30, 0.25) | *2.79 (p = 0.04)<br>-0.28 (-0.57, 0.02) |
| **Different Researchers Same Gender** | 114 | 57.7 (1.04) | -0.37 (p = 0.76)<br>0.03 (-0.25, 0.30) | | 2.42 (p = 0.07)<br>-0.26 (-0.55, 0.04) |
| **Different Researchers Different Genders** | 61 | 55.2 (1.05) | *-2.79 (p = 0.04)<br>0.28 (-0.02, 0.57) | -2.42 (p = 0.07)<br>0.26 (-.04, 0.55) | |
| **QOL-AD follow-up 2 (*NOT SIGNIFICANT IN ANALYSIS MODEL*)** | | | | | |
| **Same Researcher Same Gender** | 91 | 36.5 (0.87) | | -1.04 (p = 0.25)<br>0.12 (-0.15, 0.39) | -1.12 (p = 0.27)<br>0.15 (-0.17, 0.46) |
| **Different Researchers Same Gender** | 121 | 37.5 (0.78) | 1.04 (p = 0.25)<br>-0.12 (-0.39, 0.15) | | -0.08 (p = 0.93)<br>0.01 (-0.27, 0.30) |
| **Different Researchers Different Genders** | 65 | 37.6 (0.80) | 1.12 (p = 0.27)<br>-0.15 (-0.46, 0.17) | 0.08 (p = 0.93)<br>-0.01 (-0.30, 0.27) | |

**Significant at the 0.05 level. **Significant at the 0.01 level.*

As an interaction term between the covariate (QoL-AD baseline scores) and independent variable (gender of researcher in attendance) on the follow-up scores violated the homogeneity of regression slopes assumption and is therefore included as a term in the model, the interpretation of the results should be in relation to this interaction (Grace-Martin, 2013). Presented in Figure 15 is a scatter diagram of the QoL-AD baseline values and the follow-up 1 QoL-AD scores presented overall and split by the gender of researcher in attendance. The graph indicates that overall, the baseline scores have a positive relationship with the follow-up scores with a $R^2$ coefficient of 0.43. When split by group the relationship between baseline and follow-up scores is weaker in different researchers of

different genders group $R^2 = 0.17$ compared with the same researcher of the same gender $R^2 = 0.59$ and different groups of the same gender $R^2 = 0.32$.

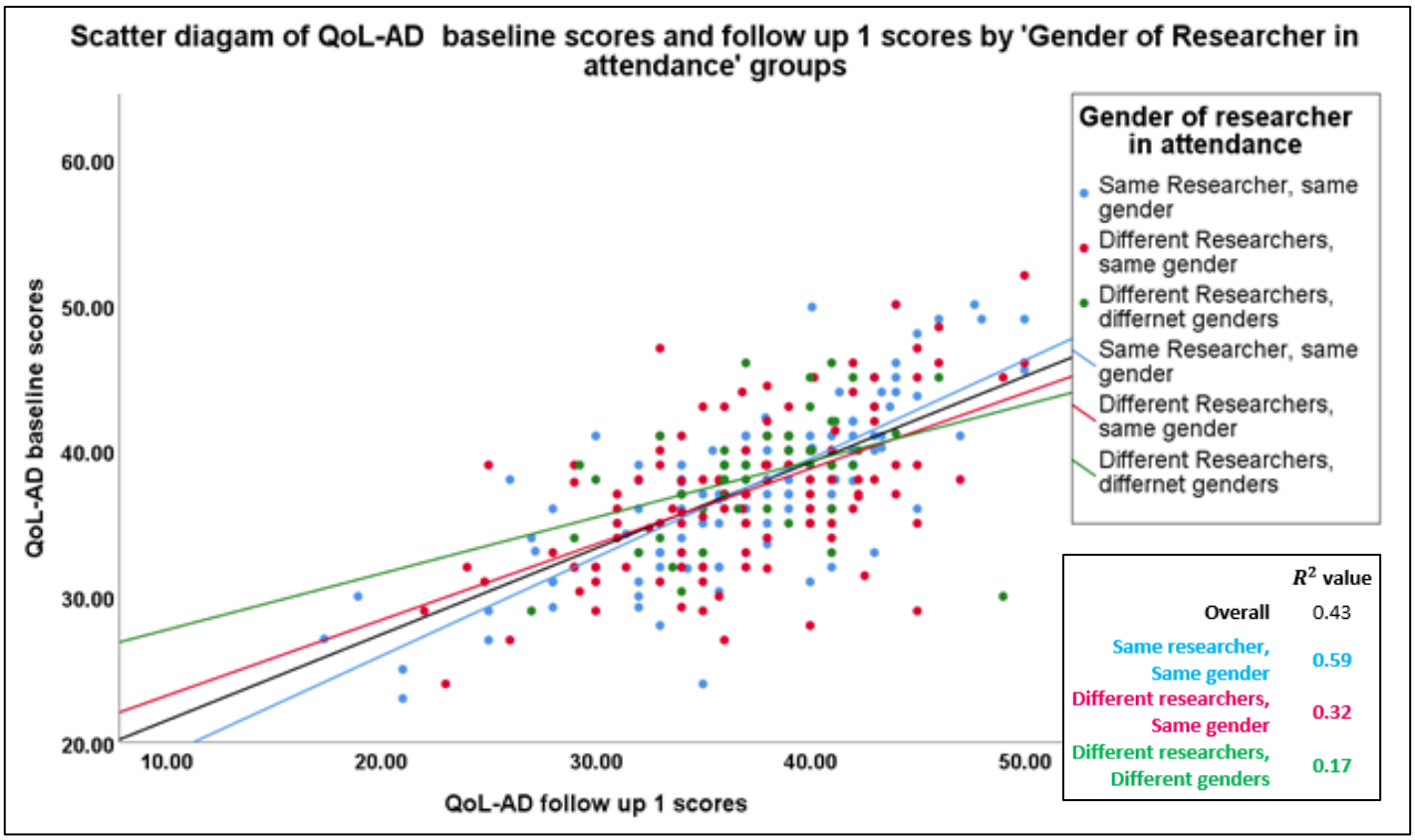


***Figure 15:*** *Scatter Plot of PwD QoL-AD Baseline and Follow-up 1, by Gender of Researcher in Attendance*

QCPR at follow-up 1

The gender of the researcher in attendance variable is statistically significant at the 5% level $F(2, 240) = 3.29$, $p = 0.04$ on the QCPR at follow-up 1. Pairwise comparisons, detailed in Table 16, were conducted and as there were no significant interaction terms between the gender of the researcher in attendance variable and covariates the effect can be interpreted from these pairwise tests. The comparisons reveal that there are statistically significant differences of the adjusted means between the different researchers of different genders group and the other two groups, same researcher same gender group ($p = 0.03$) and different researcher same gender group ($p = 0.01$). Those who had different researchers of different genders had on average higher mean QCPR scores than those with the same researcher and same gender (mean diff = 2.60) and overall higher mean scores compared to those who had different researchers of the same gender (mean diff = 3.44). The effect size of these differences are 0.26 and 0.34 respectively indicating small-medium effect sizes (Cohen, 1988).

QOL-AD at follow-up 2

At follow-up 2, the gender of researcher in attendance variable is not statistically significant at the 1% or 5% level for the QoL-AD scores. The only statistically significant finding is on the centre variable *F(*6, 240) = 2.67, p = 0.02. Post hoc tests would need to be conducted to further evaluate these differences.

QCPR at follow-up 2

On the QCPR at follow-up 2 the gender of researcher in attendance variable is statistically significant at the 5% level *F*(2, 232) = 3.90, p = 0.02. The PwD Age is also significant on the measure at the 5% level, *F*(1, 230) = 3.83, p = 0.05, indicating that PwD Age has an impact on these scores.

The pairwise comparisons between the gender of researcher in attendance groups, displayed in Table 16 above, indicate that there is a significant difference between the same researcher same gender group and the different researchers of different gender group. The Cohens D effect size calculated for this significant difference is -0.28, which indicates a small effect size (Cohen, 1988). A similar effect size of 0.26 was found between the different researcher of different genders group and the different researchers of the same gender group but is not statistically significant at the 5% level. A larger mean difference was noted between the same researcher same gender group and the different researchers of the same gender with a Cohens D effect size of 0.03 being calculated indicating a very small effect size (Cohen, 1988), again this difference is not statistically significant. However, the Cohen's effect sizes obtained are very small indicating that the mean difference is trivial (Cohen, 1988).

As the interaction term between the baseline scores and gender of researcher in attendance was identified as significant during assumption checks and therefore included in the model interpretation of these results should be guided by this (Grace-Martin, 2013). Figure 16 below displays a scatter diagram of the baseline and outcome scores split by gender of researcher in attendance groups. The diagram indicates that overall there is a positive relationship between baseline and follow up scores as expected with an $R^2$ value of 0.14. When split by gender of researcher in attendance groups this relationship is stronger overall for those in the same researcher of the same gender group ($R^2 = 0.31$). The relationship between baseline and follow up scores is weaker in the different researchers of different genders ($R^2 = 0.17$) but is weakest in the different researchers of the same gender ($R^2 = 0.04$).


***Figure 16:*** *Scatter Plot of PwD QCPR Baseline and Follow-up 2, by Gender of Researcher in Attendance*

5.2.2 Carer

The results from the carer analysis of the researcher gender attendance variable is detailed in Table 17 and the adjusted means of the gender of researcher in attendance variable are presented in Table 18.

***Table 17:*** *ANCOVA Model Analysis of Gender of Researcher in Attendance Carer Results*

| Factor | DF | F-value | Significance (p-value) | Factor | DF | F-value | Significance (p-value) |
|---|---|---|---|---|---|---|---|
| **QCPR follow-up 1** | | | | **QoL-AD proxy follow-up 1** | | | |
| **QCPR carer Baseline** | 1 | 289.91 | **<0.01 | **QoL-AD proxy Baseline** | 1 | 319.5 | **<0.01 |
| **PwD Age** | 1 | 0.16 | 0.69 | **PwD Age** | 1 | 0.73 | 0.40 |
| **Carer Gender** | 1 | 1.02 | 0.31 | **Carer Gender** | 1 | 0.18 | 0.67 |
| **Carer Age** | 1 | 0.10 | 0.75 | **Carer Age** | 1 | 0.05 | 0.83 |
| **PwD Gender** | 1 | 4.34 | *0.04 | **PwD Gender** | 1 | 0.13 | 0.72 |
| **Carer Marital status** | 1 | 4.16 | *0.04 | **Carer Marital status** | 1 | 0.60 | 0.44 |
| **Centre** | 6 | 0.76 | 0.60 | **Centre** | 6 | 0.90 | 0.50 |
| **Wave** | 4 | 1.04 | 0.39 | **Wave** | 4 | 0.97 | 0.43 |
| **Allocation** | 1 | 3.63 | 0.06 | **Allocation** | 1 | 2.00 | 0.16 |
| **Centre x Allocation** | 6 | 1.45 | 0.20 | **Centre x Allocation** | 6 | 2.15 | *0.05 |
| **Researcher Gender Attendance (fu1)** | 2 | 0.33 | 0.72 | **Researcher Gender Attendance (fu1)** | 2 | 1.94 | 0.15 |
| **Centre x Researcher Attendance** | 9 | 1.38 | 0.20 | **Centre x Researcher Attendance** | 9 | 0.80 | 0.61 |
| **Error (SS Within)** | 262 | | | **Error (SS Within)** | 277 | | |

| Factor | DF | F-value | Significance (p-value) | Factor | DF | F-value | Significance (p-value) |
|---|---|---|---|---|---|---|---|
| **QCPR follow-up 2** | | | | **QoL-AD proxy follow-up 2** | | | |
| **QCPR carer Baseline** | 1 | 188.24 | **<0.01 | **QoL-AD proxy Baseline** | 1 | 215.68 | **<0.01 |
| **PwD Age** | 1 | 0.98 | 0.32 | **PwD Age** | 1 | 0.45 | 0.50 |
| **Carer Gender** | 1 | 0.63 | 0.43 | **Carer Gender** | 1 | 3.95 | *0.05 |
| **Carer Age** | 1 | 2.03 | 0.16 | **Carer Age** | 1 | 2.10 | 0.15 |
| **PwD Gender** | 1 | 0.04 | 0.85 | **PwD Gender** | 1 | 1.35 | 0.25 |
| **Carer Marital status** | 1 | 0.01 | 0.91 | **Carer Marital status** | 1 | 0.08 | 0.78 |
| **Centre** | 6 | 1.30 | 0.26 | **Centre** | 6 | 2.12 | *0.05 |
| **Wave** | 4 | 0.32 | 0.87 | **Wave** | 4 | 1.20 | 0.31 |
| **Allocation** | 1 | 0.02 | 0.88 | **Allocation** | 1 | 0.62 | 0.43 |
| **Centre x Allocation** | 6 | 1.16 | 0.33 | **Centre x Allocation** | 6 | 0.63 | 0.70 |
| **Researcher Gender Attendance** | 2 | 0.34 | 0.71 | **Researcher Gender Attendance** | 2 | 4.29 | *0.02 |
| **Centre x Researcher Gender Attendance** | 10 | 1.23 | 0.27 | **Centre x Researcher Gender Attendance** | 10 | 1.24 | 0.27 |
| **Error (SS Within)** | 268 | | | **PwD Age x Researcher Gender Attendance** | 2 | 3.90 | *0.02 |
| | | | | **Error (SS Within)** | 278 | | |

**Significant at the 0.05 level. **Significant at the 0.01 level.*

***Table 18:*** *Estimated Marginal Means of Researcher Attendance*

| OUTOME MEASURE | ADJUSTED VALUES | | | | | |
|---|---|---|---|---|---|---|
| | Same Researcher Same gender | | Different Researchers Same gender | | Different Researchers Different Genders | |
| | N | Mean (SE) | N | Mean (SE) | N | Mean (SE) |
| **CARER QCPR FU1** | 145 | 52.5 (1.00) | 105 | 51.8 (1.11) | 47 | 52.5 (1.78) |
| **QOL-AD PROXY FU1** | 153 | 31.9 (0.68) | 112 | 30.5 (0.76) | 47 | 31.6 (1.23) |
| **CARER QCPR FU2** | 113 | 53.3 (1.47) | 125 | 52.2 (1.28) | 66 | 51.9 (1.92) |
| **QOL-AD PROXY FU2** | 114 | 31.8 (0.88) | 134 | 30.4 (0.76) | 68 | 31.2 (1.15) |

**POST HOC PAIRWISE COMPARISONS FOR QOL-AD PROXY MEASURE AT FOLLOW-UP 2**

| | MEAN DIFFERENCES (p-value of difference) COHEN'S D EFFECT SIZE (95% CI) | | |
|---|---|---|---|
| | Same Researcher Same gender | Different Researchers Same gender | Different Researchers Different Genders |
| **Same Researcher Same Gender** | | 1.34 (p = 0.11)<br>-0.15 (-0.40, 0.10) | 0.61 (p = 0.62)<br>-0.06 (-0.35, 0.23) |
| **Different Researchers Same Gender** | -1.34 (p = 0.11)<br>0.15 (-0.10, 0.40) | | -0.72 (p = 0.55)<br>0.09 (-0.19, 0.36) |
| **Different Researchers Different Genders** | -0.61 (p = 0.62)<br>0.06 (-0.23, 0.35) | 0.72 (p = 0.55)<br>-0.09 (-0.36, 0.19) | |

QoL-AD proxy at follow-up 1

The gender of researcher in attendance variable is not statistically significant on either outcome measures at follow-up 1. However, the interaction between centre and allocation is statistically significant at the 5% level $F(6, 277) = 2.15$, $p = 0.05$ on the QoL-AD proxy. To evaluate where these differences lie, post hoc pairwise comparison tests would need to be conducted.

QCPR at follow-up 1

The gender of researcher in attendance variable is also not statistically significant on the QCPR measure at follow-up 1. Variables which are statistically significant at the 5% level on the QCPR at follow-up 1 include the PwD gender $F(1, 262) = 4.34$, $p = 0.04$ and carer marital status $F(1, 262) = 4.16$, $p = 0.04$.

QCPR at follow-up 2

There are no statistically significant results on the carer QCPR measure at follow-up 2 including the gender of researcher in attendance variable.

QOL-AD at follow-up 2

The carer gender is statistically significant at the 5% level $F(1, 278) = 3.95$, $p = 0.05$, on the QoL-AD proxy at follow-up 2 along with a statistically significant result at the 5% level on the Centre variable $F(3, 273) = 3.09$, $p = 0.03$. The gender of researcher in attendance variable is also significant at the 5% level $F(2, 278) = 4.29$, $p = 0.02$ on the measure. Pairwise comparisons of the researcher gender in attendance are displayed in table 18. The results indicate that the differences between the measures at each level of the variable are not statistically significant at the 1% or 5% level and the Cohen D effect sizes calculated for the differences (ranging from 0.06 to 0.15) are small. Because the mean differences and Cohen's effect sizes obtained were very small, this indicates that the mean difference is trivial but there is an effect of the variable present in the model (Cohen, 1988).

Since the interaction term between the covariate (PwD Age) and independent variable is included in the model the interpretation should consider this (Grace-Martin, 2013). Figure 17 contains a scatter diagram of PwD Age variable against the QoL-AD proxy follow-up 2 scores presented overall and split by gender of researcher in attendance levels. The graph indicates that, overall, the age of the participant has a very weak negative relationship with the QoL-AD proxy follow-up 2 scores with a $R^2$ coefficient of 0.02. When split by the gender of researcher in attendance groups this relationship is slightly stronger for those who had the same researcher of the same genders ($R^2 = 0.07$). This relationship is weaker for those who had different researchers of different genders ($R^2 = 0.015$) and is weakest in the different researchers of the same genders ($R^2 = 0.002$).

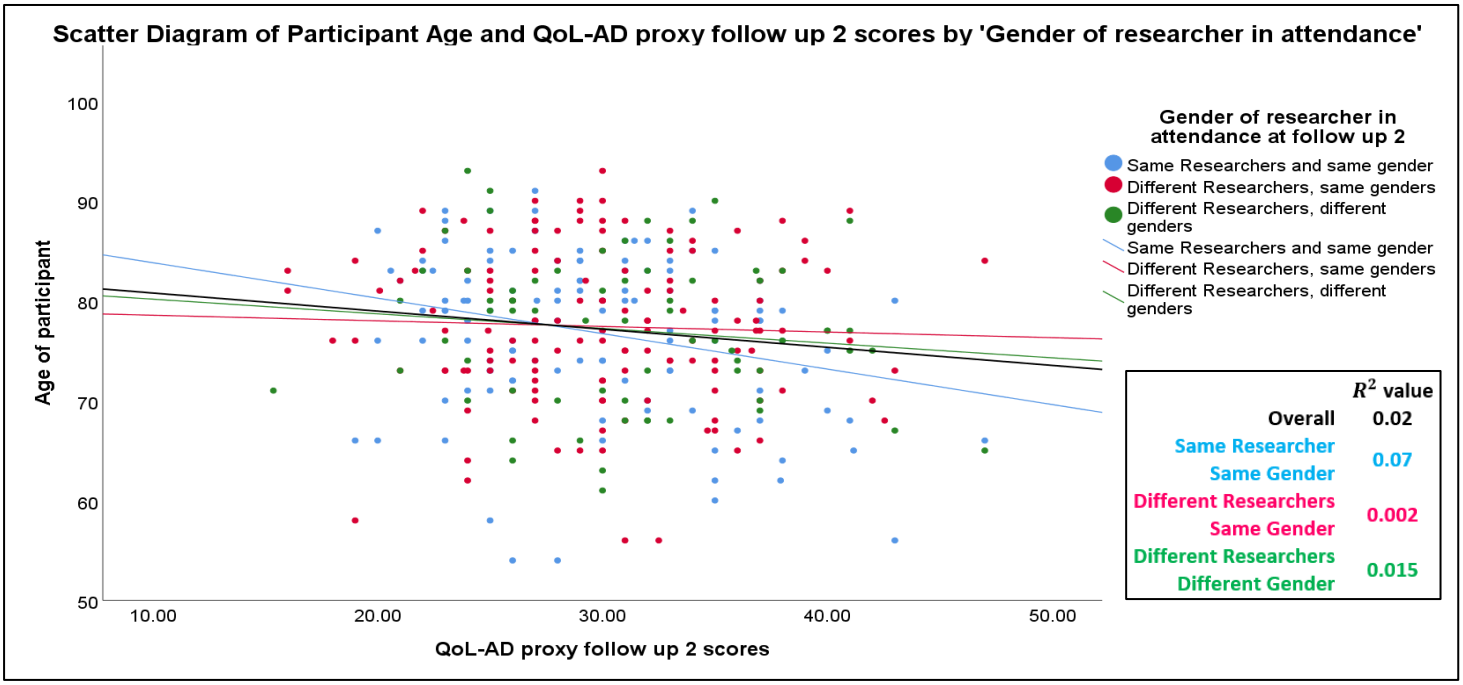


***Figure 17:*** *Scatter Plot of PwD Age and Carer QoL-AD Follow-up 2, by Gender of Researcher in Attendance*

## 5.3 Sensitivity analysis

As a few of the researchers' genders could not be assumed from the name, four unknown researcher gender cases were removed in the main analysis models. Sensitivity analysis has been conducted including these four unknown genders - firstly as male researchers and then subsequently as female researchers to evaluate whether there is an impact on the analysis results of these known. Results from these sensitivity analyses are contained in Tables 1 through to 4 of Appendix 11.

### **Participant** QoL-AD at follow-up 1

The gender of researcher in attendance variable was significant on the QoL-AD following primary analysis at the 1% level of significance $F(2, 258) = 5.74$, $p = 0.01$. The variable was still statistically significant on the outcome measure in the sensitivity analysis models when assuming the unknown researchers to be female $F(2, 262) = 5.48$, $p < 0.01$ and then male $F(2, 262) = 5.42$, $p < 0.01$.

### **Participant** QCPR at follow-up 1

The sensitivity analysis on the QCPR at follow-up 1 revealed that, as in primary analysis, the gender of researcher in attendance variable is statistically significant at the 5% level when assuming these unknown researcher genders as female and then males $F(2, 242) = 3.20$, $p = 0.05$ and $F(2, 242) = 3.16$, $p = 0.04$ respectively.

The interaction between baseline scores and the gender of researcher in attendance variable on the QCPR follow-up scores was not statistically significant during primary analysis but was significant when assuming the unknown researcher genders to be both female and male. The sensitivity results also indicated a discrepancy on the Wave variable. For primary analysis the wave variable was not statistically significant but is statistically significant at the 5% level on both analysis for assuming female and then male $F(4, 242) = 2.46$, $p = 0.05$ and $F(4, 242) = 2.49$, $p = 0.04$ respectively

### **Participant** QOL-AD and QCPR at follow-up 2

On the QCPR at follow-up 2 the gender of researcher in attendance variable is not statistically significant at the 1% or 5% level at primary or sensitivity analyses however is significant on the QoL-AD measure at the 5% level ($p = 0.02$). The significance of the variable remains the same on both measures in all three analysis models, either when removing the unknown researcher gender cases (primary analysis) or when assuming the unknowns to be female or male (sensitivity analysis).

A discrepancy was noted between primary analysis and sensitivity analysis on the centre variable on the QoL-AD measure at follow-up 2. During primary analysis Centre was statistically significant at the 5% level $F(6, 244) = 2.26$, $p = 0.04$ and is also statistically significant at the 5% level $F(6, 247) = 2.38$, $p = 0.03$ in the sensitivity analysis results when assuming the unknown researcher genders to be female. However, when assuming the unknowns to be male the variable is not statistically significant at the 5% or 1% level $F(6, 246) = 2.08$, $p = 0.06$.

**Carer proxy QOL-AD and QCPR at follow-up 1**

As with the participant data, the QoL-AD proxy and QCPR carer models at follow-up 1 produced consistent results of the gender of researcher in attendance variable. When removing unknown researcher gender cases, assuming unknown to be male or assuming unknown to be female the model results remain the same with there being statistically significant effects at the 5% or 1% levels.

The allocation on the QCPR follow-up 1 produced a result in the primary analysis which was not statistically significant at the 5% level, $F(1, 262) = 3.36$, $p = 0.06$ yet close to the 0.05 significance threshold. In both sensitivity models when assuming either female or male the allocation variable was now *just* significant at the 5% level $F(1, 266) = 3.89$, $p = 0.05$ and $F(1, 266) = 3.96$, $p = 0.05$ respectively. The discrepancies are not large inconsistencies and whilst they technically alter the statistical significance of the variable the results of all three models are around the boundary of statistical significance (i.e. p value less than 0.05) and the sensitivity analysis produces no major inconsistencies.

There is also a similar finding on the interaction between centre and allocation on the QoL-AD at follow-up 1. On the primary analysis the interaction is statistically significant at the 5% level, $F(6, 277) = 2.15$, $p = 0.05$ and also when assuming the unknowns to be male $F(6, 281) = 2.22$, $p = 0.04$. However the interaction is not significant at the 5% level when assuming the unknown researcher genders to the female $F(6, 281) = 2.05$, $p = 0.06$. Again, the results of the main analysis model and the sensitivity analysis models are close to the boundary of statistical significance (i.e. p value 0.05) and whilst the results do alter the statistical significance, they are very small discrepancies.

**Carer proxy QOL-AD and QCPR at follow-up 2**

Again, the QoL-AD proxy and QCPR carer models at follow-up 2 were consistent across the three analyses models for the results of the gender of researcher in attendance variable with no

significant findings at the 1% or 5% level on the QCPR measure but all three being significant at the 5% level on the QoL-AD proxy measure.

On the QoL-AD measure, several variables produced inconsistent results in terms of statistical significance across the three analysis models. Centre was significant at the 5% on the primary analysis $F(6, 278) = 2.12$, $p = 0.05$ and was significant at the 1% level $F(6, 276) = 3.32$, $p = 0.01$ the unknown researchers to be male, $F(6, 277) = 1.93$, $p = 0.08$. However, when assuming the unknown researchers to be female the centre variable is not statistically significant at the 5% level.

The carer gender factor also produced different results between primary and sensitivity analysis. When assuming the unknowns to be female or male the factor is not statistically significant in the model at the 1% or 5% level however when removing the unknown genders from analysis in the primary model the factor is statistically significant at the 5% level $F(1, 278) = 3.95$, $p = 0.05$. Additionally, the interaction between centre and gender of researcher in attendance produced discrepancies between the models. For the primary analysis and sensitivity analysis when assuming the unknown genders to be female the factor is not statistically significant in either models at the 1% or 5% significance levels. However, when assuming the unknown researchers to be male the factor is statistically significant at the 5% level $F(10, 282) = 2.10$, $p = 0.03$.

### Sensitivity analysis summary

The aim of the sensitivity analysis was to assess whether the four cases of unknown researcher genders being removed from the analysis data set had an impact on the results. Having analysed these unknowns firstly as female and then as male and comparing the results, it appears that the unknowns have not had an impact on the gender of researcher in attendance variable and therefore does not require further investigation. Some of the other variables in the models were affected, however, these results vary, and are not the primary concern of the current analysis. As there a very few cases of unknown researcher gender this impact on the variables is somewhat unexpected, but the variables concerned tend to have small samples representing groups and therefore results of these should be treated with caution. The inconsistencies with the p-value between these analyses are very small and hover around the boundary of statistical significance (i.e. p value 0.05).

## 5.4 Chapter Summary

The current chapter evaluated whether the gender of the attending researcher impacts upon the outcome measure scores by conducting ANCOVA models to evaluate any differences between the

means of the levels of the gender of researcher in attendance variable. The gender of researcher in attendance variable is significant on both the participant outcomes at follow-up 1, at the 5% level on the QCPR and the 1% level on the QoL-AD and is also significant at the 5% level on the QCPR measure at follow-up 2. Table 19 below summarises the main findings of the gender of research in attendance analysis.

***Table 19**: Summary of results of analysis for gender of researcher in attendance variable*

| | Outcome | Gender of Researcher in Attendance variable |
|---|---|---|
| **Participant** | **QCPR follow up 1** | Significant at 5% level (p=0.04) |
| | **QoL-AD follow up 1** | Significant at 1% level (p=0.01) |
| | **QCPR follow up 2** | Significant at 5% level (p=0.02) |
| | **QoL-AD follow up 2** | Not Significant (p=0.26) |
| **Carer** | **QCPR follow up 1** | Not Significant (p=0.72) |
| | **QoL-AD follow up 1** | Not Significant (p=0.15) |
| | **QCPR follow up 2** | Not Significant (p=0.71) |
| | **QoL-AD follow up 2** | Significant at 5% level (p=0.02) |

Post hoc tests on the QCPR at follow-up 1 and follow-up 2 indicated that those who had 'different researchers of different genders' had on average higher mean QCPR scores than those with the 'same researcher of the same gender' and overall higher mean scores compared to those who had 'different researchers of the same gender'. These differences produced small-medium effect sizes.

The post hoc tests on the participant QoL-AD at follow-up 1 indicated that the relationship between baseline and follow-up scores is stronger in the same researcher and same gender group and less strong/prominent in the different researchers group. When looking at the two groups who had different researchers, those who had different researchers of the same gender had a stronger relationship between baseline and follow-up than those with researchers of different genders.

The carer data is not statistically significant at the 5% level on the QCPR and proxy QoL-AD at follow-up 1 nor on the QCPR at follow-up 2 but is significant on the proxy QoL-AD at follow-up 2. Post hoc tests revealed that overall the age of the participant has a very weak negative relationship with the QoL-AD proxy follow-up 2 scores. When split by the gender of researcher in attendance groups this relationship is slightly stronger for those who had the same researcher of the same gender but is weaker for those who had different researchers of the same gender or different genders.

# Chapter Six – Discussion and Conclusion

The current research study aimed to explore two research questions, in a study with three time points for persons with Dementia (PwD) and carer dyads:

1. Does the use of a different researcher at time points impact upon the outcome measure?
2. Does the gender of the attending researcher impact upon the outcome measure?

Data from a previously conducted RCT, the REMCARE study, was used where researchers attended visits to collect data at baseline, follow-up 1 (3 months after baseline) and follow-up 2 (10 months after baseline (7 months after follow-up 1)). The outcome measures explored were two quality of life (QoL) measures the quality of carer patient relationship (QCPR) and the quality of life Alzheimers Disease (QoL-AD) (Logsdon et al., 1999; Spruytte et al., 2002). The QCPR evaluates the quality of the relationship between the participant and carer from both their perspectives and the QoL-AD measures QoL specifically in Alzheimer's disease but is widely used in all Dementia populations. The final data sets used contained 330 participants in each totalling 660 (330 dyads - the PwD and their corresponding carer). At follow-up 1 160 participants had the same researcher and 170 had different researchers. At follow-up 2, 118 had the same researcher for all visits, 129 had two different researchers across visits and 83 had all three different researchers. For the gender of researcher in attendance 160 had the same researcher therefore the same gender, 119 had different researchers of the same gender and 47 had different researchers with a representative of both genders. At follow-up 2 118 had the same researcher and therefore the same gender, 139 had different researchers all the same gender and 69 had different researchers with a representative of both genders.

Several underlying hypotheses contributed to the research questions. Firstly, it is hypothesised that if the continuity of researcher attendance or gender of researcher in attendance impacts upon the outcome measures, then you would expect to see a statistically significant effect of the variables in the participant analysis models at the follow-ups. The researcher attendance effect is thought to impact upon the outcome measures collected, as participants build up a relationship and rapport with the researcher during subsequent visits. This may affect the nature of the responses provided by the PwD/carer dyad and the direction of affect in subsequent follow-ups as researchers had prior knowledge of the participant's state or responses at baseline. As a result, it was hypothesised that the effect may be stronger at follow-up 2 due to researcher continuity.

The underlying hypothesis for the second research question is that gender may play a role in rapport building and may impact upon participant responses (Pollner, 1988; Thurnell-Read, 2016; Williams & Heikes, 1993). Therefore, as the gender of the researcher in attendance varies, we would expect to see an impact in the *post hoc* analysis. As previously stated in Chapter Three, it should be noted that the issue of gender fluidity goes beyond the scope of this thesis and here when referring to researcher gender it was related to the researcher's biological sex at birth based on the researcher's name. The analysis was therefore subject to the limitations which are described in more detail below.

Another underlying hypothesis, with both variables (researcher attendance and gender of researcher in attendance) is that you would not expect to see an effect on the carer completed outcome measures. This is because the carers completed questionnaires independently with little interaction and little assistance from the researcher. As a result, if researcher attendance or gender bias is present, an effect should not be seen in the carer data. It was also hypothesised that you would expect the effect to be larger on the QCPR measure compared with the QoL-AD outcome measure. This is because the QCPR measure contains more 'personal' questions regarding the relationship of the participant and carer. Here, the researcher is required to discuss with participants their personal relationships with the carer, which may be difficult for participants to disclose, especially if the rapport between researcher and patient is poor. Therefore, rapport between researcher and participant is expected to have more of an impact on this measure in comparison.

To explore the research questions, ANCOVA models were run on both measures (QCPR and QoL-AD) at follow-up 1 and follow-up 2 on the two datasets - participant and carer. The models evaluated if there are any statistically significant effects of the researcher attendance variable and the gender of researcher in attendance variable on the outcome measures at the corresponding time point indicating whether or not there are any statistical differences of the mean outcome measure scores between the variable groups. The main results and interpretations are discussed below.

## Main Results

The primary analysis on the participant data set revealed a statistically significant finding on the researcher attendance variable at follow-up 1 on both the QCPR and QoL-AD outcome measures indicating that there is a difference between mean scores of the same researcher group and different researcher group. There was also a statistically significant finding seen on the interaction between researcher attendance variable and centre on the QCPR measure at follow-up 2. The primary analysis at follow-up 2 for the participant data did not produce statistically significant findings on either the

QCPR or the QoL-AD measure main effects indicating no difference between the means of the three researcher groups (same, two different, three different). Sensitivity analysis was conducted on the follow-up 2 measures where participants who had a break in continuity of researcher at the second visit were re-categorised into the 'three different researchers' group as opposed to being in the 'two different' group. The results of this analysis found a statistically significant effect of the researcher attendance variable on both the QCPR and Qol-AD measures indicating a difference between the means at follow-up 2 when re-grouped. These re-grouping results indicate that the break in continuity has an impact on the results which contributes to the evidence that there is a researcher effect.

The results on researcher gender found similar patterns. A statistically significant effect was seen on both the QCPR and QoL-AD measures at follow-up 1. This suggests that there was a difference in the means scores between the levels of the researcher gender attendance groups; same researcher of the same gender, different researchers of the same gender and different researchers of different genders. At follow-up 2, there were no statistically significant findings on the researcher gender attendance variable on the QoL-AD, however, the variable was statistically significant on the QCPR measure.

The results of *post hoc* tests on the follow-up 1 models indicate that on the QCPR, the effect is largest between those who had different researchers of the same gender and those who had different researchers of different genders. There was also a statistically significant effect between those who had the same researcher and same gender and those who had different researchers of a different gender. There was not a statistically significant difference between those who had the same and those who had different researchers of the same gender. At follow-up 2 the significant finding between the same researcher of the same gender and different researchers of different genders was present but the statistically significant finding between the different researchers of the same gender and different researchers of different genders was no longer present.

Similarly, for the QoL-AD measure at follow-up 1 the relationship between baseline and follow-up scores is strongest in the same researcher same gender group and weakest in the different researchers of different genders. The different researchers of the same gender group had a stronger relationship between baseline and follow-up scores than the different researcher of different genders but a weaker relationship than the same researchers of the same gender. Finding a difference between these groups could suggest that there is some impact of the gender of the researcher in

attendance and not just the researcher attendance, however, this is very difficult to disentangle and interpret. Additionally, as mentioned in the sections below these results aren't definitive.

Also, in relation to both research questions it was hypothesised that as the carers had little interaction with the researchers and completed questionnaires independently then the 'researcher effect' would not be present on the carer outcome measures. The results of the carer models revealed that the researcher attendance variable and the gender of researcher in attendance variable was not statistically significant on the QCPR at follow-up 1, the QoL-AD at follow-up 1 nor on the QCPR at follow-up 2 which supports this hypothesis and contributes to the evidence indicating a researcher effect. However, in contrast to this, the QoL-AD at follow-up 2 did find a statistically significant finding on both the researcher attendance variable and the gender of researcher in attendance variable which contradicts this hypothesis. For the researcher attendance variable, as with the participant follow-up 2 measures, sensitivity analysis was conducted on the carer follow-up 2 outcomes following the re-grouping of participants who had a continuity break at the middle time point (i.e. those who had a different researcher at the middle visit). The results of this sensitivity analysis on the carer outcomes did not differ from the primary analysis indicating that the re-grouping had no effect, again supporting the hypothesis that the carer measures are not affected by the researcher in attendance.

It is also thought that the researcher bias may be more prominent on the QCPR measure compared with the QoL-AD outcome measure as the QCPR has more in-depth personal questions. It is therefore hypothesised that the effect will be larger on the QCPR measure in comparison. However, the results of the researcher attendance analysis at follow-up 1 may indicate that the effect size is slightly larger on the QoL-AD measure contradicting this hypothesis. Follow-up 2 primary analysis revealed no significant findings therefore effect sizes were not obtained. Cohens D effect sizes were calculated for the sensitivity analysis significant results at follow-up 2. The magnitude of these effect sizes also indicates in general that the QoL-AD produced higher effects suggesting that the personal questions may not have had more of an impact. However, the analysis of the gender of researcher in attendance variable revealed the opposite. Larger Cohens D effect sizes were obtained at follow-up 1 on the QCPR measure compared to the effect sizes obtained for the QoL-AD measure. At follow-up 2 these are not comparable as the QoL-AD was not statistically significant in the model.

### Results interpretation and explanation

The theory under-pinning the hypotheses for this study was that the relationship between participants and individual researchers is built over time (Pitts & Miller-Day, 2007). As a result, we

would expect an effect at follow-up 2, if an effect was present at follow-up 1. However, our primary analysis results on both the researcher attendance variable and the gender of researcher in attendance variable show that there was an effect at Follow-Up 1 (in QCPR and QoL-AD), but not at follow-up 2.

This could be related to on-going cognitive decline in the PwD study population. In general, Dementia patients' symptoms decline over time and although each case of Dementia is different, the loss of short-term memory is a common symptom (Association, 2013; "How Dementia progresses", n.d.; Wu et al., 2016). It is possible that at the second follow-up appointment, participants may not be able to recall the researcher they saw at baseline or follow-up 1 and any rapport built up in previous assessments may have been lost. In addition, the time period between the follow-up assessments may add to this. Baseline data was collected at baseline and follow-up 1 was three months after this, the final follow-up was not until a further seven months, meaning that there is a much larger gap between follow-up 1 and follow-up 2 compared with the first two visits. Therefore, the longer timeframe between follow-up 1 and follow-up 2 has more potential to allow patients to 'forget' their researcher and allows for this rapport being diminished along with it being more likely that the symptoms of cognitive decline are more prominent as the study and disease progress. Similarly, if the researcher effect seen was due to the researcher having prior knowledge of the participant's state and previous responses this longer time frame may have had an impact on the researcher's recall.

From a statistical perspective, the way the variables are constructed could contribute to the effect not being seen at follow-up 2 but being present at follow-up 1. There are more groups at follow-up 2 compared with follow-up 1 for both factors of interest; researcher attendance variable (two at follow-up 1, three at follow-up 2) and gender of researcher in attendance variable (three at follow-up 1 and four at follow-up 2). This means that smaller samples are represented in each group at follow-up 2 resulting in less statistical power and the chance of making a Type 1 error is higher; therefore, it is more likely at follow-up 2 that, as power is diminished, an effect may be present but not seen.

Another explanation for the different results at follow-up 1 and follow-up 2 could be in relation to the way in which the variable is constructed and an anomaly caused by how the data is being coded. This appears to be supported by the results of the sensitivity analysis. Re-coding the researcher attendance variable at follow-up 2 (where patients in the 'two different researchers group' who had a different researcher conduct their middle (follow-up 1) assessment were re-categorised into the three different researchers group as they represented a break in continuity)), produced a

statistically significant finding, with the effect being stronger at follow-up 2, when compared to follow-up 1. This would be in line with the under-pinning hypothesis that rapport builds over time and explain the results of the primary analysis. Such an approach to the analysis could be justified, as it can be argued that any development of rapport would be interrupted, should participants be seen by a different researcher in the interim visit. There is a potential argument that although short term memory is affected and the time points are spaced the hypothesized continuity and build-up of rapport through visits is still having an impact and is important to consider.

The researcher effect hypothesis was expected to be seen in the participant data and not the carer data. Whilst most of the results support this hypothesis the QoL-AD measure produced a significant effect on the researcher attendance variable and gender of researcher in attendance variables at follow-up 2 in the models. This suggests that although the carer completes the measure independently, they may still be influenced by researcher continuity; although the study aimed for carers to complete the questionnaires independently with limited interaction, they may have had some input from the researcher. Equally, carers may have been influenced by researcher continuity as the measure used requires the carer to comment on their PwD, which could be influenced by researcher rapport. However, several elements could have contributed to this result which cannot be measured here and the impact of researcher consistency on carer data and reliability of the results should be further explored using a different data-set.

It was also hypothesised that the researcher effect would be larger in the QCPR outcome compared with the QoL-AD measure, as the QCPR measure includes more personal questions regarding the relationship between participant and carer. Findings of the post hoc tests conducted reveal that larger effect sizes were seen on the QoL-AD measure for the researcher attendance variable, which does not support this. However, this may be caused by the nature of the personal questions in both measures and whether they were sensitive enough to detect any differences caused by researcher rapport.

In contrast, the analysis of the gender of researcher in attendance found larger effect sizes at follow-up 1 on the QCPR compared to the QoL-AD. This indicates that researcher gender may be influential. This concurs with a number of studies in the literature, which appear to suggest that participants are more likely to disclose information to female interviewers and reports of sensitive topics are more likely to be noted by female researchers regardless of the gender of the subject being interviewed (Pollner, 1988). In this study most researchers were female so this cannot be evaluated,

however, it could contribute to the bigger effect being seen on the QCPR measure in the current results. It should be noted that the QoL-AD outcome measure is well established and has shown good reliability in several studies (Bowling et al., 2015; Ettema, et al., 2005; Selai & Trimble, 1999). In contrast, the reliability of the QCPR is yet to be substantiated and this may have contributed to the differences in effects seen on the measures.

Chapter Two briefly explored the converse idea of 'too much rapport' (Hill & Hall, 1963; Miller, 1952), which could also influence how researchers record their scores on the key measures used. It could be that where participants have a good relationship with the same researcher, the participants may find it harder to respond to personal questions. The phenomena of 'desirability and faking good' and conversely 'deviation and faking bad' discussed in Chapter One could apply here. The direction of the hypothesis used in this thesis assumes that a positive rapport between participants and researchers is built over time and leads to more accurate recording of outcome measures. It may be the case that participants have the same researcher but have a bad rapport with them and so find it difficult to disclose sensitive information. This analysis was not aimed to measure the extent or direction of rapport between researcher and participant and the findings here are not conclusive in either direction. The results merely indicate whether there is a difference (or not) caused by researcher continuity.

## Comparisons with literature

To the author's knowledge, the research questions explored here have not been addressed elsewhere and research in this area is sparse. One other study has explored the impact of researcher continuity on outcome measure collection (Kobak, 2010). It hypothesised that "researcher bias" may be more prevalent in those situations where the same researcher collects both baseline and follow-up measures. The current study findings contribute to this argument as the results indicate that the relationship between baseline and follow-up scores are stronger in the same researcher group. This means that those who had lower baseline scores are more likely to have lower follow-up scores and vice versa. Whilst this is expected in QoL research, this appears to be more prominent in cases where the same researcher undertakes both assessments (Hoe et al., 2009; Gräske et al., 2018).

As with previous studies in other disease areas, such as pain and sexual health, the current results indicate that there is an impact of researcher gender on outcome measures which could be a result of researcher gender influencing participant responses unintentionally (Fisher, 2007; McClelland & McCubbin, 2008; Miyazaki & Taylor, 2008). However, whilst the current results

demonstrate a possible effect between researcher gender and attendance, they do not indicate which gender is more influential and the direction of effect.

## Statistical Approach and Methods

The analysis methods used in this thesis aligned to those that were undertaken for the REMCARE study. Adopting the same model and factors enabled the research to introduce additional variables (researcher attendance and gender of researcher in attendance variables) in order to evaluate the effects that they produced. The REMCARE study conducted both an ANCOVA model and a Linear Mixed Model (LMM) at the two follow-ups on the primary and secondary outcomes. Given the two research questions under investigation, it was important to evaluate changes at specific time points. As a result, ANCOVA models were preferred to LMMs. LMMs would have allowed an analysis across time, but as the dataset only had two follow-up time points that were unevenly spaced, LMMs would have been more difficult to interpret.

In terms of the results of the second research question regarding the gender of the researcher in attendance. There were limitations on what could be analysed statistically. As stated in Chapter Two, some researchers suggests that good rapport can be built more easily between researchers and participants of the same gender (Williams & Heikes, 1993), however, the interaction between researcher and participant gender has not been looked at here. The sample size available would not have enough study power to look at this interaction and results would be limited.

## Study Limitations

In addition to the limitations highlighted in the preceding section, the current research had other limitations caused by the trial from which the data set was drawn. The analyses presented here were explorative in nature and REMCARE was not designed to test these hypotheses. Firstly, randomisation was not used to allocate participants to researcher attendances groups. Randomisation is used to allocate participants to a group based on certain important characteristics. This means that confounding variables were not accounted for (Akobeng, 2005). Although the splits between researcher attendance groups are relatively similar (see Chapter Three Table 1), randomisation has not been used to control each strata level within groups.

Equally, the researchers partaking in the study were not randomised or controlled for. Many researchers (43 in total) were used and the variability between these researchers was not calculated or accounted for and the reliability of these individuals on scoring the measures is unknown. Several

centres were used from different areas across the UK and there is a large amount of variation in the number of participants between centres and between the researcher attendance groups. For example, Bangor site had few researchers and therefore many of their subjects had the same researcher; in contrast, the larger sites (Manchester and London) used researcher networks with many different researchers conducting follow-ups. Centre effects were included in the analysis models but as there was not an even number of researcher attendance groups, this means that centre effects might not have been adequately accounted for. Outcome measure scoring could vary by centre and/or researcher and the experience and training of each individual researcher was also not taken into account (given the lack of data). This may have contributed to researcher variation across the study and affected results.

Another aspect of the study design, which should not be overlooked, is statistical power. Any well designed RCT would have conducted a sample size (power calculation) *a priori* and aimed to recruit the required number of participants derived in this calculation in order to minimise Type II error (stating an effect isn't there when it is). This thesis used a subset of data from the original REMCARE study. Because of the research question proposed, I only included data where participants and carers had completed all three visits during the study. Therefore, the power of the current research was not calculated *a priori* and there is a possibility that a Type II error may have been present.

Another limitation of this research was the assumptions made when coding certain variables to be used for analysis. Firstly, the coding of the two independent variables of interest the 'researcher attendance' and 'gender of researcher in attendance'. Participants were placed in 'researcher attendance' groups based on the researcher name detailed on the 'completed by' variable. This assumes that the researcher listed in the variable was the researcher who completed the entire assessment and CRF and it is unknown whether the researcher listed was the researcher who went out to conduct the entire assessment. In some cases, the assessment may have been incomplete and finished by another researcher at a different time point. Equally, it is possible that two researchers conducted these visits but only one was listed in the CRF, which could have distorted the results.

Other assumptions were also made on this 'completed by' variable during the coding. In cases where two researchers were listed, the first was assumed to be the 'main' researcher and groupings were based on this. Whilst this may have caused an incorrect allocation, there was only one case noted which should have little impact on the results. However, there may have been other cases where two

researchers attended together but only one listed. There were also several assumptions made on certain researcher names. As described in Chapter Three, where the researcher last name was given along with the full first name or initial of first name (i.e. 'Rachel Evans' and 'R Evans') they were assumed the same. Minor spelling mistakes were assumed to be the same researcher (i.e. 'Rachel' Evans and 'Raechl Evans'. Whilst the assumptions made were minor, the coding of same or different researcher based on these assumptions may have affected the allocation.

When coding patients into 'gender of researcher in attendance group' a few other assumptions were made. The researcher gender was determined based on the researcher name in the 'completed by' variable. A website was used which allowed you to enter a researcher name and produced probabilities for gender based on the name entered ("GenderChecker.com", n.d). The reliability and accuracy of this website is unknown and its reliability is considered to be somewhat lower than if relying on source data or study CRFs. Errors may have been made when inferring the researcher gender from the name which could have had an impact on the groupings and again on the study results. As stated, when referring to gender throughout this thesis it has been using the assumption of sex at birth and has not considered gender fluidity. Categorising researchers into male or female alone does not necessarily evaluate the full impact of researcher 'gender' characteristics on participant responses. Researchers may have characteristics or traits which could influence participant responses and rapport building that are not necessarily attributable to their gender or biological sex. In addition, we have not considered researcher age, training, experience, warmth, ability to establish rapport or any other variables in relation to researcher demographics or characteristics which could all have an impact. Furthermore, as stated in the statistical approach, we have merely explored the gender of researcher in attendance groups and not considered the impact of the gender of the participant along with the researcher gender. The interaction between PwD gender and researcher gender should be explored to evaluate impacts of researcher gender on outcome measures in further depth. This study however did not have a large enough sample size to look at this.

Other variables that were used in the analysis were also coded manually that may have impacts on the data and results. The number of visits each patient had from the original full REMCARE data was coded based on the 'completed by' variable and placed into categories of either 1, 2 or 3 visits. As this variable was based on whether or not the completed by variable was completed with a researcher name and nothing else the accuracy of the number of visits may be unreliable. Some subjects from the sample may have been said to not have completed a particular visit if this data was

not present but they might have actually completed the assessment. This may have affected sample size and study power of the current analysis as some subjects may have been excluded.

On the other hand, the completed by variable may have been present for a subject even though the visit was not conducted meaning that some subjects included in the sample should not have been. At follow-up 1, as the analysis was run as complete case, any missing data would have been excluded and this would not be an issue; similarly, at follow-up 2 if the outcome was missing this would have been removed from analysis. However, if at follow-up 2 the follow-up 1 score was missing then this subject should not have been included in analysis. The outcome measure completion rates are very high in the sample at each time point therefore this is unlikely to be a common occurrence and should not impact the data or results greatly but again should be noted as a possible limitation.

Another limitation, which stems from the original REMCARE data, is with regards to researcher blinding. The REMCARE study design aimed to have the researchers collecting outcome data blinded to the participant allocation, however due to the nature of the intervention participants were unblind. This resulted in the possibility of researchers becoming accidently unblinded during the follow-up assessments through the patients giving them information that indicated their allocation. As blinding of researchers is essential to reducing bias of outcome measure collection (Akobeng, 2005; Hoare, 2010) this means that potentially the data collected may have been influenced by this. For the REMCARE study to monitor this, researchers completed perception sheets following a patient assessment indicating to which treatment group they believed the patient was allocated. Results from analysing these perception sheets indicated that researchers 'were indeed more likely to be correct than incorrect in the direction of their prediction' (Woods et al., 2012). Therefore, possible bias introduced during the outcome measure collection should be considered as a limitation of the current findings.

Although the study protocol intended for carers to complete assessments alone with little input from the researcher, some researchers assisted some carers where required - yet this was not recorded and not accounted for in this analysis. Therefore, where it is assumed no researcher interaction in the carer data, this may not be true, and the researcher may have influenced carer outcomes even though results indicate no researcher effect.

Due to the above limitations highlighted, as with any research, the potential for bias is present and results should be interpreted with these limitations in mind.

## Recommendations for future research

This explorative study indicates a possible researcher attendance and gender effect on the outcome measure scores of participants in a Dementia trial. Whilst the results imply that there is a possible effect present, they are not definitive and subject to several limitations as outlined above. It should be noted that the study was not designed nor powered to detect this signal and all results and findings should be interpreted with caution. Future trials should be conducted to establish causality and future research studies on this topic should be carefully and specifically designed. In agreement with, it is suggested that research looking to explore this topic should use randomisation to assign patients to researcher attendance or gender in attendance groups to establish this question more definitively (Kobak, 2010). Independent research studies could be conducted and specifically designed to assess this topic. A study within a study (SWAT) which is a "self-contained research study that has been embedded within a host trial with the aim of evaluating or exploring alternative ways of delivering or organising a particular trial process" could be used (Treweek et al., 2018). The main intention of a SWAT is to evaluate methods used to conduct a trial process to provide evidence about how to improve the process (Treweek et al., 2018).

It should be considered whether Dementia studies are the most appropriate disease area to evaluate this topic of research. Dementia commonly causes issues with cognitive functioning and tends to affect short term memory (Association, 2013; Wu et al., 2016). One of the underlying assumptions underpinning researcher continuity is that rapport is built over time. This may be more difficult to establish in Dementia trials due to the patient's short-term memory. However, researcher consistency may also have an impact on outcome measure based on researchers influence of previous knowledge of the participant. As a result, this impact could/should still be explored. In addition, different outcome measures should be explored to evaluate which are impacted by researcher continuity as some may be more sensitive than others.

If an effect is demonstrated through these future studies then implications for the conduct and design of future RCTs would need consideration. RCTs could incorporate these effect in one of two ways - either through the logistical aspects of the research design or by incorporating some researcher effect into the analysis model.

RCTs could be designed so that researchers collecting outcome data attend visits to the same participant where possible. This may be very difficult for larger studies with several sites and would depend on the size and structure of the involved sites. Smaller areas, such as for example in this study

Bangor or Gwent, may be able to accommodate this for data collection as few researchers work at the site. However, as larger sites such as London or Manchester may have multiple researchers, conducting follow-ups and organising the same researchers to attend visits for the same participants could be logistically difficult, especially for the sites that utilise clinical research nurse (CRN) networks. Aspects such as staff turn around, number of follow-ups, follow-up length and number of patients would determine whether or not this could be achieved. A team from Sheffield CTU (Clinical Trials Unit) working on a Dementia trial 'Journeying through Dementia (JtD)' attempted to do this in their study in order to improve retention rates (Wright et al., 2019). The team introduced continuity of researchers part way through the study to increase retention rates. They stated that this took a lot of effort but felt it was integral to its success and worth doing. At the time of this thesis, they had no official statistics on the success or uptake of their approach, but believe continuity improved after it was implemented. Issues that occurred were staff turn around and staff becoming unblinded. Therefore, organising researcher consistency may be difficult and constraining.

Another way in which this researcher effect could be accounted for is by including the factor in the analysis model to control for the researcher groups (same, different or mixture) variation. This would have some implications for research design as it would need to be ensured that data to include this in the model is collected and that the researcher groups are represented in all comparison arms.

Whilst it may be logistically difficult to ensure research data is collected by the same or different researchers this may further reduce bias in research and improve the reliability of outcome measures collected which is a huge importance to trials. The attempted efforts to minimise this bias should depend on how large the effect is, if established. Researchers, trial methodologists and clinicians would need to assess the impact of using different researchers in their study designs against the logistical implications on the trial and should take action where necessary and appropriate.

Evidence based medicine (EBM) relies on research to inform clinical practice. To form this body of evidence it is essential to evaluate the effectiveness of the intervention and RCTs are acknowledged as the 'gold standard' of research methods to do this (Brocklehurst et al., 2019; Moore et al., 2015; Treweek et al., 2018). Due to their importance it is imperative that RCTs are conducted to the highest standards, yet, a considerable amount of trials are regarded as low quality (Brocklehurst et al., 2019; Treweek et al., 2018). Ironically, whilst EBM focuses on using evidence to inform decisions for health care, trials themselves are conducted using processes with little evidence to support their use (Treweek et al., 2018). A lot of work into improving trial quality is being conducted, for example

through Trial Forge (Trial Forge, n.d.), which is “an initiative that aims to increase the evidence base for trial decision making”. They implement the use of SWATs to grow this body of evidence and improve trial efficiency (Treweek et al., 2018). The focus of these SWATs is largely around trial recruitment and retention as both are critical to the success of a clinical trial and achieve sufficient statistical power. However, as stated in Chapter Two reliability of the outcome measure is also considered important for statistical power and there is “growing recognition that insufficient attention has been paid to the outcomes measured in clinical trials” (Kirkham et al., 2016).

Any observed effect in clinical trials is based on this outcome measurement, therefore the choice, collection and evaluation of the outcome measure are critical for the evaluation of interventions in trials. Many trials that are adequately powered show no effect of the intervention even though researchers and investigators are confident the Interventions work, for example, the REMCARE study (Kobak, 2010; Woods et al., 2012). This raises the question of whether there are other aspects relating to trial design and conduct influencing this outcome measure such as a researcher effect or bias discussed here. It could potentially be due to the use of the wrong outcome measure or the design of measures such as question format or response options etc., however many measures used in trials have been evaluated to demonstrate good reliability and validity (Bowling et al., 2015; Ready & Ott, 2003; Soobiah et al., 2019). One useful way to explore unexpected trial results is by conducting a process evaluation alongside studies (Craig et al., 2008).

The importance of conducting a process evaluation alongside clinical trials for complex interventions is becoming more and more acknowledged as they can assist with the explanation of trial results (Brocklehurst et al., 2019; Moore et al., 2015). If the intervention is successful, they can help assess the mechanisms of how and why it was successful along with which components of the intervention were important (Craig et al., 2008; Public Health England, 2018). In contrast, if the intervention is found to be unsuccessful in improving health outcomes then process evaluations can help unpack potential reasons for why the intervention was ineffective (Craig et al., 2008). They may also be used to help explain and determine why an intervention may work for some populations, certain settings and specific contexts but are not impactful in other settings (Public Health England, 2018). Process evaluations can be conducted alongside RCTS or independently and generally adopt qualitative methods but can also use quantitative techniques (Public Health England, 2018).

In addition, process evaluations can assist with obtaining knowledge regarding the trial conduct and design. They can collect data on the perceptions of the study from the participants,

researchers, clinicians, investigators and other personnel involved (Oakley et al., 2006). Future investigation in the area of researcher continuity and researcher gender attendance should consider conducting a process evaluation in parallel to quantitative assessments as it could contribute to investigation of any researcher effects or impacts regarding outcome measure collection in more detail. They could assist in starting to unpack the mechanisms of any researcher continuity effect as to whether any bias stems from the researcher prior knowledge of the participant and/or from the influence of the researcher on participant responses and could aid interpretation.

## Implications for practice

Implications for practice from this study are very limited and, as stated, further research is required to establish any researcher continuity effect definitively as this study was never designed to definitively detect this signal. As stated above, arguably Dementia trials may not the most appropriate example to best test these hypotheses and this population may not be impacted however it could be further explored.

## Conclusion

This study aimed to use data from a previously conducted RCT with three time points to evaluate for persons with Dementia (PwD) and carer dyads firstly, 'does the use of a different researcher at time points impact upon the outcome measure?' and secondly, 'does the gender of the attending researcher impact upon the outcome measure?' Table 20 below summarises the overall contribution of this thesis to the research.

***Table 20**: Summary of what this thesis adds*

| |
|---|
| **Background and Literature** |
| ➢ The importance of good quality RCTs for EBM is highly recognised |
| ➢ Outlined bias in trials with a focus on outcome measure reliability |
| ➢ Raises the question of whether the use of multiple rater's has an impact on outcome measurement in Dementia Trials |
| **Research questions** |
| in a study with three time points for persons with Dementia (PwD) and carer dyads:<br>1. Does the use of a different researcher at time points impact upon the outcome measure?<br>2. Does the gender of the attending researcher impact upon the outcome measure? |
| **Main Results** |
| ➢ Results show an indication of an effect of researcher continuity within this data set |
| ***Researcher Attendance analysis*** |
| ➢ Variable significant on both outcomes at follow up 1 on the participant data |
| ➢ Follow up 2 primary analysis found no significant effect |
| ➢ Sensitivity analysis at follow up 2 produced significant findings on both outcomes |
| ➢ Carer QoL-AD proxy measure significant at follow up 2 |
| ➢ Carer sensitivity analysis produced same results as main analysis |
| ***Gender of Researcher in Attendance analysis*** |
| ➢ Variable significant on both participant outcomes at follow up 1 and on the participant QPCR at follow up 2 |
| ➢ Carer QoL-AD proxy measure significant at follow up 2 |
| **Future Research in this area** |
| ➢ Further investigation into the effect of these variables is required |
| ➢ Future studies should be designed a priory |
| ➢ SWATs could be utilised, and a parallel process evaluation conducted |
| **Implications for practice** |
| ➢ If effect is established through further research the design of future RCTs may need to consider this bias |

For the PwD a statistically significant effect of the researcher attendance variable and the gender of researcher in attendance variable was found on both outcome measures at follow-up 1 but not at follow-up 2. For the carer data, analysis revealed no statistically significant effects on either measures at follow-up 1 for the researcher attendance and gender of researcher attendance variables however both variables were found to be statistically significant on the QoL-AD measure at follow-up 2 for the carer.

The results from this research indicate a possible impact of researcher continuity and/or researcher gender continuity however the study was not designed to establish any causal effects nor was it powered to; therefore, results should be treated with caution. Further research should be conducted in this area to establish any impact of researcher collection on outcome measures in RCTs and if found the design of future RCTs may need to consider this bias.

## Bibliography


A. Kobak, K. (2010). Inaccuracy in Clinical Trials: Effects and Methods to Control Inaccuracy. *Current Alzheimer Research*, *7*(7), 637–641. https://doi.org/10.2174/156720510793499057

*A Word About Evidence: 6. Bias - a proposed definition*. (2018, June 15). Catalog of Bias. https://catalogofbias.org/2018/06/15/a-word-about-evidence-6-bias-a-proposed-definition/

Addington-Hall, J., & Kalra, L. (2001). Who should measure quality of life? *BMJ (Clinical Research Ed.)*, *322*(7299), 1417–1420. PubMed. https://doi.org/10.1136/bmj.322.7299.1417

Akobeng, A. K. (2005). Understanding randomised controlled trials. *Archives of Disease in Childhood*, *90*(8), 840–844. https://doi.org/10.1136/adc.2004.058222

Alexopoulos, G. S., Abrams, R. C., Young, R. C., & Shamoian, C. A. (1988). Cornell Scale for Depression in Dementia. *Biol Psychiatry*, *23*, 271–284.

Allen, M. (2002). Randomized Clinical Trials: Attention to Bias. *Journal of Wound, Ostomy and Continence Nursing*, *29*(3), 127–128.

Alzheimers Society. (n.d.). *Dementia UK update. Second Edition*. Retrieved 29 March 2020, from https://www.who.int/news-room/fact-sheets/detail/dementia

American Psychological Association. (2012). Guidelines for the evaluation of dementia and age-related cognitive change. *American Psychologist*, *67*(1), 1–9. https://doi.org/10.1037/a0024643

Andrade, C. (2015). The Primary Outcome Measure and Its Importance in Clinical Trials: (Clinical and Practical Psychopharmacology). *The Journal of Clinical Psychiatry*, *76*(10), e1320–e1323. https://doi.org/10.4088/JCP.15f10377

Arnold, D. M. (2011). Bias in transfusion research: From study design to result reporting. *Transfusion*, *51*(2), 237–240. https://doi.org/10.1111/j.1537-2995.2010.02990.x

Association, A. P. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Publishing.

Banerjee, S. (2006). Quality of life in dementia: More than just cognition. An analysis of associations with quality of life in dementia. *Journal of Neurology, Neurosurgery & Psychiatry*, *77*(2), 146–148. https://doi.org/10.1136/jnnp.2005.072983

Banerjee, Sube, Samsi, K., Petrie, C. D., Alvir, J., Treglia, M., Schwam, E. M., & del Valle, M. (2009). What do we know about quality of life in dementia? A review of the emerging evidence on the predictive and explanatory value of disease specific measures of health related quality of life in people with dementia. *International Journal of Geriatric Psychiatry*, *24*(1), 15–24. https://doi.org/10.1002/gps.2090

Beerens, H. C., Zwakhalen, S. M. G., Verbeek, H., Ruwaard, D., & Hamers, J. P. H. (2013). Factors associated with quality of life of people with dementia in long-term care facilities: A systematic review. *International Journal of Nursing Studies*, *50*(9), 1259–1270. https://doi.org/10.1016/j.ijnurstu.2013.02.005

Bell, K., Fahmy, E., & Gordon, D. (2016). Quantitative conversations: The importance of developing rapport in standardised interviewing. *Quality & Quantity*, *50*(1), 193–212. https://doi.org/10.1007/s11135-014-0144-2

Bialocerkowski, A. E., & Bragge, P. (2008). Measurement error and reliability testing: Application to rehabilitation. *International Journal of Therapy and Rehabilitation*, *15*(10), 422–427. https://doi.org/10.12968/ijtr.2008.15.10.31210

Black, B. S., Johnston, D., Morrison, A., Rabins, P. V., Lyketsos, C. G., & Samus, Q. M. (2012). Quality of life of community-residing persons with dementia based on self-rated and caregiver-rated measures. *Quality of Life Research*, *21*(8), 1379–1389. https://doi.org/10.1007/s11136-011-0044-z

Bosboom, P. R., Alfonso, H., Eaton, J., & Almeida, O. P. (2012). Quality of life in Alzheimer's disease: Different factors associated with complementary ratings by patients and family carers. *International Psychogeriatrics*, *24*(5), 708–721. https://doi.org/10.1017/S1041610211002493

Bothwell, L. E., & Podolsky, S. H. (2016). The Emergence of the Randomized, Controlled Trial. *New England Journal of Medicine*, *375*(6), 501–504. https://doi.org/10.1056/NEJMp1604635

Bowling, A., Rowe, G., Adams, S., Sands, P., Samsi, K., Crane, M., Joly, L., & Manthorpe, J. (2015). Quality of life in dementia: A systematically conducted narrative review of dementia-specific measurement scales. *Aging & Mental Health*, *19*(1), 13–31. https://doi.org/10.1080/13607863.2014.915923

Bremer, P., Cabrera, E., Leino-Kilpi, H., Lethin, C., Saks, K., Sutcliffe, C., Soto, M., Zwakhalen, S. M. G., & Wübker, A. (2015). Informal dementia care: Consequences for caregivers' health and health care use in 8 European countries. *Health Policy*, *119*(11), 1459–1471. https://doi.org/10.1016/j.healthpol.2015.09.014

Brocklehurst, P., & Hoare, Z. (2017). How to design a randomised controlled trial. *British Dental Journal*, *222*(9), 721–726. https://doi.org/10.1038/sj.bdj.2017.411

Brocklehurst, P. R., Baker, S. R., Listl, S., Peres, M. A., Tsakos, G., & Rycroft-Malone, J. (2019). How Should We Evaluate and Use Evidence to Improve Population Oral Health? *Dental Clinics of North America*, *63*(1), 145–156. https://doi.org/10.1016/j.cden.2018.08.009

Brod, M., Stewart, A. L., Sands, L., & Walton, P. (1999). Conceptualization and Measurement of Quality of Life in Dementia: The Dementia Quality of Life Instrument (DQoL). *The Gerontologist*, *39*(1), 25–36. https://doi.org/10.1093/geront/39.1.25

Cerejeira, J., Lagarto, L., & Mukaetova-Ladinska, E. B. (2012). Behavioral and Psychological Symptoms of Dementia. *Frontiers in Neurology*, *3*. https://doi.org/10.3389/fneur.2012.00073

Charter, R. A. (1999). Sample Size Requirements for Precise Estimates of Reliability, Generalizability, and Validity Coefficients. *Journal of Clinical and Experimental Neuropsychology*, *21*(4), 559–566. https://doi.org/10.1076/jcen.21.4.559.889

Chenoweth, L., & Jeon, Y.-H. (2007). Determining the efficacy of Dementia Care Mapping as an outcome measure and a process for change: A pilot study. *Aging & Mental Health*, *11*(3), 237–245. https://doi.org/10.1080/13607860600844226

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). L. Erlbaum Associates.

Connor, D. J., & Sabbagh, M. N. (2008). Administration and Scoring Variance on the ADAS-Cog. *Journal of Alzheimer's Disease*, *15*(3), 461–464. https://doi.org/10.3233/JAD-2008-15312

Cook, C., & Sheets, C. (2011). Clinical equipoise and personal equipoise: Two necessary ingredients for reducing bias in manual therapy trials. *Journal of Manual & Manipulative Therapy*, *19*(1), 55–57. https://doi.org/10.1179/106698111X12899036752014

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: The new Medical Research Council guidance. *BMJ*, a1655. https://doi.org/10.1136/bmj.a1655

Crofton, J., & Mitchison, D. A. (1948). Streptomycin Resistance in Pulmonary Tuberculosis. *British Medical Journal*, *2*(4588), 1009–1015.

Davis, R. E., Couper, M. P., Janz, N. K., Caldwell, C. H., & Resnicow, K. (2010). Interviewer effects in public health surveys. *Health Education Research*, *25*(1), 14–26. https://doi.org/10.1093/her/cyp046

DeVon, H. A., Block, M. E., Moyle-Wright, P., Ernst, D. M., Hayden, S. J., Lazzara, D. J., Savoy, S. M., & Kostas-Polston, E. (2007). A Psychometric Toolbox for Testing Validity and Reliability. *Journal of Nursing Scholarship*, *39*(2), 155–164. https://doi.org/10.1111/j.1547-5069.2007.00161.x

*Diagnostic and statistical manual of mental disorders: DSM-5™, 5th ed.* (pp. xliv, 947). (2013). American Psychiatric Publishing, Inc. https://doi.org/10.1176/appi.books.9780890425596

Doody, O., & Noonan, M. (2013). Preparing and conducting interviews to collect data. *Nurse Researcher*, *20*(5), 28–32. https://doi.org/10.7748/nr2013.05.20.5.28.e327

Dow, J., Robinson, J., Robalino, S., Finch, T., McColl, E., & Robinson, L. (2018). How best to assess quality of life in informal carers of people with dementia; A systematic review of existing outcome measures. *PLOS ONE*, *13*(3), e0193398. https://doi.org/10.1371/journal.pone.0193398

Ettema, T. P., Dröes, R.-M., de Lange, J., Mellenbergh, G. J., & Ribbe, M. W. (2007). QUALIDEM: Development and evaluation of a dementia specific quality of life instrument. Scalability, reliability and internal structure. *International Journal of Geriatric Psychiatry*, *22*(6), 549–556. https://doi.org/10.1002/gps.1713

Ettema, T. P., Droes, R.-M., Lange, J. de, Mellenbergh, G. J., & Ribbe, M. W. (2005). A review of quality of life instruments used in dementia. *Quality of Life Research*, *14*(3), 675–686. https://doi.org/10.1007/s11136-004-1258-0

*Facts for the media*. (n.d.). Alzheimer's Society. Retrieved 29 March 2020, from https://www.alzheimers.org.uk/about-us/news-and-media/facts-media

Fisher, R. A. (1992). Statistical Methods for Research Workers. In S. Kotz & N. L. Johnson (Eds.), *Breakthroughs in Statistics: Methodology and Distribution* (pp. 66–70). Springer New York. https://doi.org/10.1007/978-1-4612-4380-9_6

Fisher, T. D. (2007). Sex of Experimenter and Social Norm Effects on Reports of Sexual Behavior in Young Men and Women. *Archives of Sexual Behavior*, *36*(1), 89–100. https://doi.org/10.1007/s10508-006-9094-7

Fossey, J., Lee, L., & Ballard, C. (2002). Dementia care mapping as a research tool for measuring quality of life in care settings: Psychometric properties. *International Journal of Geriatric Psychiatry*, *17*(11), 1064–1070. https://doi.org/10.1002/gps.708

Frost, M. H., Reeve, B. B., Liepa, A. M., Stauffer, J. W., & Hays, R. D. (2007). What Is Sufficient Evidence for the Reliability and Validity of Patient-Reported Outcome Measures? *Value in Health*, *10*, S94–S105. https://doi.org/10.1111/j.1524-4733.2007.00272.x

Gaur, R., Bullock, R., Nath, G., Kakati, R., & De Santi, S. (2009). P1-218: Are we adequately evaluating and monitoring rater performance in clinical trails with dementia? *Alzheimer's & Dementia*, *5*(4S_Part_8), P240–P240. https://doi.org/10.1016/j.jalz.2009.04.225

*GenderChecker.com*. (n.d.). [Gender Checker]. Retrieved 29 March 2020, from https://genderchecker.com/

Gerhard, T. (2008). Bias: Considerations for research practice. *American Journal of Health-System Pharmacy*, *65*(22), 2159–2168. https://doi.org/10.2146/ajhp070369

Gluud, L. L. (2006). Bias in Clinical Intervention Research. *American Journal of Epidemiology*, *163*(6), 493–501. https://doi.org/10.1093/aje/kwj069

Grace-Martin, K. (2013, December 22). *ANCOVA Assumptions: When Slopes are Unequal*. The Analysis Factor. https://www.theanalysisfactor.com/ancova-assumptions-when-slopes-are-unequal/

Gräske, J., Schmidt, A., Schmidt, S., Laporte Uribe, F., Thyrian, J. R., Michalowsky, B., Schäfer-Walkmann, S., & Wolf-Ostermann, K. (2018). Quality of life in persons with dementia using regional dementia care network services in Germany: A one-year follow-up study. *Health and Quality of Life Outcomes*, *16*(1), 181. https://doi.org/10.1186/s12955-018-0990-z

Guillemin, M., & Heggen, K. (2009). Rapport and respect: Negotiating ethical relations between researcher and participant. *Medicine, Health Care and Philosophy*, *12*(3), 291–299. https://doi.org/10.1007/s11019-008-9165-8

Gwet, K. L. (2014). *Handbook of Inter-Rater Reliability, 4th Edition: The Definitive Guide to Measuring The Extent of Agreement Among Raters*. Advanced Analytics, LLC.

Harrison, J. K., Noel-Storr, A. H., Demeyere, N., Reynish, E. L., & Quinn, T. J. (2016). Outcomes measures in a decade of dementia and mild cognitive impairment trials. *Alzheimer's Research & Therapy*, *8*(1), 48. https://doi.org/10.1186/s13195-016-0216-8

*Health and social care professionals*. (n.d.). Alzheimer's Society. Retrieved 29 March 2020, from https://www.alzheimers.org.uk/get-support/help-dementia-care/health-and-social-care-professionals

Heinze, G., & Dunkler, D. (2017). Five myths about variable selection. *Transplant International*, *30*(1), 6–10. https://doi.org/10.1111/tri.12895

Hey, S. P., London, A. J., Weijer, C., Rid, A., & Miller, F. (2017). Is the concept of clinical equipoise still relevant to research? *BMJ*, j5787. https://doi.org/10.1136/bmj.j5787

Higgins, J. P. T., Altman, D. G., Gotzsche, P. C., Juni, P., Moher, D., Oxman, A. D., Savovic, J., Schulz, K. F., Weeks, L., Sterne, J. A. C., Cochrane Bias Methods Group, & Cochrane Statistical Methods Group. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, *343*(oct18 2), d5928–d5928. https://doi.org/10.1136/bmj.d5928

Higgins J.P.T, Altman D.G., Sterne JAC (editors). Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Churchill R, Chandler J, Cumpston MS (editors), Cochrane Handbook for Systematic Reviews of Interventions version 5.2.0 (updated June 2017), Cochrane, 2017. Available from www.training.cochrane.org/handbook.

Hill, R. J., & Hall, N. E. (1963). A Note on Rapport and the Quality of Interview Data. *The Southwestern Social Science Quarterly*, *44*(3), 247–255.

Hoare, Z. (2010). Randomisation: What, why and how? *Significance*, *7*(3), 136–138. https://doi.org/10.1111/j.1740-9713.2010.00443.x

Hoe, J., Hancock, G., Livingston, G., Woods, B., Challis, D., & Orrell, M. (2009). Changes in the Quality of Life of People With Dementia Living in Care Homes: *Alzheimer Disease & Associated Disorders*, *23*(3), 285–290. https://doi.org/10.1097/WAD.0b013e318194fc1e

*How dementia progresses*. (n.d.). Alzheimer's Society. Retrieved 31 March 2020, from https://www.alzheimers.org.uk/about-dementia/symptoms-and-diagnosis/how-dementia-progresses

Huang, H.-L., Chang, M. Y., Tang, J. S.-H., Chiu, Y.-C., & Weng, L.-C. (2009). Determinants of the discrepancy in patient- and caregiver-rated quality of life for persons with dementia. *Journal of Clinical Nursing*, *18*(22), 3107–3118. https://doi.org/10.1111/j.1365-2702.2008.02537.x

IBM Corp. Released 2017. *IBM SPSS Statistics for Windows*, Version 25.0. Armonk, NY: IBM Corp.

Indrayan, A. (2012). *Basic Methods of Medical Research* (4th edition). Aitbs publishers.

Jopling, K. (2017). *Promissing approaches to living well with Demenita*. Age UK. https://www.ageuk.org.uk/globalassets/age-uk/documents/reports-and-publications/reports-and-briefings/health--

wellbeing/rb_feb2018_promising_approaches_to_living_well_with_dementia_report.pdf

JPND research. (2015). *Dementia Outcome Measures: Charting New Territory*. JPND research. https://www.neurodegenerationresearch.eu/wp-content/uploads/2015/10/JPND-Report-Fountain.pdf

Juni, P. (2001). Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ*, *323*(7303), 42–46. https://doi.org/10.1136/bmj.323.7303.42

Kirkham, J. J., Gorst, S., Altman, D. G., Blazeby, J. M., Clarke, M., Devane, D., Gargon, E., Moher, D., Schmitt, J., Tugwell, P., Tunis, S., & Williamson, P. R. (2016). Core Outcome Set–STAndards for Reporting: The COS-STAR Statement. *PLOS Medicine*, *13*(10), e1002148. https://doi.org/10.1371/journal.pmed.1002148

Kraemer, H. C. (1979). Ramifications of a population model for κ as a coefficient of reliability. *Psychometrika*, *44*(4), 461–472. https://doi.org/10.1007/BF02296208

Lawton, M. P. (1983). The varieties of wellbeing. *Experimental Aging Research*, *9*(2), 65–72. https://doi.org/10.1080/03610738308258427

Leon, A. C. (2007). Implications of Clinical Trial Design on Sample Size Requirements. *Schizophrenia Bulletin*, *34*(4), 664–669. https://doi.org/10.1093/schbul/sbn035

Leon, Andrew C., Marzuk, P. M., & Laura, P. (1995). More Reliable Outcome Measures Can Reduce Sample Size Requirements. *Archives of General Psychiatry*, *52*(10), 867–871. https://doi.org/10.1001/archpsyc.1995.03950220077014

Lewis, F., Schaffer, S. K., Sussex, J., O'Neill, P., & Cockcroft, L. (2014). *The Trajectory of Dementia in the UK - Making a Difference*. 62.

Lindsay, B. (2004). Randomized controlled trials of socially complex nursing interventions: Creating bias and unreliability? *Journal of Advanced Nursing*, *45*(1), 84–94. https://doi.org/10.1046/j.1365-2648.2003.02864.x

Logsdon, R. G., Gibbons, L. E., McCurry, S. M., & Teri, L. (1999). Quality of life in Alzheimer's disease: Patient and caregiver reports. *Journal of Mental Health and Aging*, *5*(1), 21–32.

Logsdon, R. G., Gibbons, L. E., McCurry, S. M., & Teri, L. (2002). Assessing Quality of Life in Older Adults With Cognitive Impairment. *Psychosomatic Medicine*, *64*(3). https://journals.lww.com/psychosomaticmedicine/Fulltext/2002/05000/Assessing_Quality_of_Life_in_Older_Adults_With.16.aspx

Logsdon, R. G., & Teri, L. (1997). The Pleasant Events Schedule-AD: Psychometric Properties and Relationship to Depression and Cognition in Alzheimer's Disease Patients1. *The Gerontologist*, *37*(1), 40–45. https://doi.org/10.1093/geront/37.1.40

Lohr, K. N. (2002). Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research*, *11*(3), 193–205. https://doi.org/10.1023/A:1015291021312

Lyketsos, C. G., Gonzales-Salvador, T., Chin, J. J., Baker, A., Black, B., & Rabins, P. (2003). A follow-up study of change in quality of life among persons with dementia residing in a long-term care facility. *International Journal of Geriatric Psychiatry*, *18*(4), 275–281. https://doi.org/10.1002/gps.796

Magaziner, J., Zimmerman, S. I., Gruber-Baldini, A. L., Hebel, J. R., & Fox, K. M. (1997). Proxy Reporting in Five Areas of Functional Status: Comparison with Self-Reports and Observations of Performance. *American Journal of Epidemiology*, *146*(5), 418–428. https://doi.org/10.1093/oxfordjournals.aje.a009295

Malone, H., Nicholl, H., & Tracey, C. (2014). Awareness and minimisation of systematic bias in research. *British Journal of Nursing*, *23*(5), 279–282. https://doi.org/10.12968/bjon.2014.23.5.279

Marques, M. J., Woods, B., Hopper, L., Jelley, H., Irving, K., Kerpershoek, L., Meyer, G., Bieber, A., Stephan, A., Sköldunger, A., Sjölund, B., Selbaek, G., Rosvik, J., Zanetti, O., Portolani, E., Vugt, M., Verhey, F., Gonçalves-Pereira, M., & on behalf of the Actifcare Consortium. (2019). Relationship quality and sense of coherence in dementia: Results of a European cohort study. *International Journal of Geriatric Psychiatry*, *34*(5), 745–755. https://doi.org/10.1002/gps.5082

Marshall, I. J., Kuiper, J., & Wallace, B. C. (2015). Automating Risk of Bias Assessment for Clinical Trials. *IEEE Journal of Biomedical and Health Informatics*, *19*(4), 1406–1412. https://doi.org/10.1109/JBHI.2015.2431314

Marshall, S., Haywood, K., & Fitzpatrick, R. (2006). Impact of patient-reported outcome measures on routine practice: A structured review. *Journal of Evaluation in Clinical Practice*, *12*(5), 559–568. https://doi.org/10.1111/j.1365-2753.2006.00650.x

McClelland, L. E., & McCubbin, J. A. (2008). Social influence and pain response in women and men. *Journal of Behavioral Medicine*, *31*(5), 413–420. https://doi.org/10.1007/s10865-008-9163-6

McDonald, S., Westby, M., Clarke, M., & Lefebvre, C. (2002). Number and size of randomized trials reported in general health care journals from 1948 to 1997. *International Journal of Epidemiology*, *31*(1), 125–127. https://doi.org/10.1093/ije/31.1.125

Medvedev, O. N., & Landhuis, C. E. (2018). Exploring constructs of well-being, happiness and quality of life. *PeerJ*, *6*. https://doi.org/10.7717/peerj.4903

Merlo, P., Devita, M., Mandelli, A., Rusconi, M. L., Taddeucci, R., Terzi, A., Arosio, G., Bellati, M., Gavazzeni, M., & Mondini, S. (2018). Alzheimer Café: An approach focused on Alzheimer's patients but with remarkable values on the quality of life of their caregivers. *Aging Clinical and Experimental Research*, *30*(7), 767–774. https://doi.org/10.1007/s40520-017-0844-2

Merriam-Webster. (n.d.). Dementia. In Merriam-Webster.com dictionary. Retrieved March 30, 2020, from https://www.merriam-webster.com/dictionary/dementia

Microsoft Corporation. (2018). Microsoft Excel. Retrieved from https://office.microsoft.com/excel

Millenaar, J., Hvidsten, L., de Vugt, M. E., Engedal, K., Selbæk, G., Wyller, T. B., Johannessen, A., Haugen, P. K., Bakker, C., van Vliet, D., Koopmans, R. T. C. M., Verhey, F. R. J., & Kersten, H. (2017). Determinants of quality of life in young onset dementia – results from a European multicenter assessment. *Aging & Mental Health*, *21*(1), 24–30. https://doi.org/10.1080/13607863.2016.1232369

Miller, S. M. (1952). The Participant Observer and 'Over-Rapport'. *American Sociological Review*,

*17*(1), 97. https://doi.org/10.2307/2088368

Missotten, P., Squelard, G., Ylieff, M., Di Notte, D., Paquay, L., De Lepeleire, J., Buntinx, F., & Fontaine, O. (2008). Relationship between Quality of Life and Cognitive Decline in Dementia. *Dementia and Geriatric Cognitive Disorders*, *25*(6), 564–572. https://doi.org/10.1159/000137689

Miyazaki, A. D., & Taylor, K. A. (2008). Researcher Interaction Biases and Business Ethics Research: Respondent Reactions to Researcher Characteristics. *Journal of Business Ethics*, *81*(4), 779–795. https://doi.org/10.1007/s10551-007-9547-5

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research*, *19*(4), 539–549. https://doi.org/10.1007/s11136-010-9606-8

Moniz-Cook, E., Hart, C., Woods, B., Whitaker, C., James, I., Russell, I., Edwards, R. T., Hilton, A., Orrell, M., Campion, P., Stokes, G., Jones, R. S., Bird, M., Poland, F., & Manthorpe, J. (2017). Challenge Demcare: Management of challenging behaviour in dementia at home and in care homes – development, evaluation and implementation of an online individualised intervention for care homes; and a cohort study of specialist community mental health care for families. *Programme Grants for Applied Research*, *5*(15), 1–290. https://doi.org/10.3310/pgfar05150

Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O'Cathain, A., Tinati, T., Wight, D., & Baird, J. (2015). Process evaluation of complex interventions: Medical Research Council guidance. *BMJ*, *350*(mar19 6), h1258–h1258. https://doi.org/10.1136/bmj.h1258

Mougias, A. A., Politis, A., Lyketsos, C. G., & Mavreas, V. G. (2011). Quality of life in dementia patients in Athens, Greece: Predictive factors and the role of caregiver-related factors.

*International Psychogeriatrics*, *23*(3), 395–403. https://doi.org/10.1017/S1041610210001262

Mulsant, B. H., Kastango, K. B., Rosen, J., Stone, R. A., Mazumdar, S., & Pollock, B. G. (2002). Interrater Reliability in Clinical Trials of Depressive Disorders. *American Journal of Psychiatry*, *159*(9), 1598–1600. https://doi.org/10.1176/appi.ajp.159.9.1598

Murad, M. H., Asi, N., Alsawas, M., & Alahdab, F. (2016). New evidence pyramid. *Evidence Based Medicine*, *21*(4), 125–127. https://doi.org/10.1136/ebmed-2016-110401

Naglie, G. (2007). Quality of Life in Dementia. *Canadian Journal of Neurological Sciences / Journal Canadien Des Sciences Neurologiques*, *34*(S1), S57–S61. https://doi.org/10.1017/S0317167100005588

Naglie, G., Tomlinson, G., Tansey, C., Irvine, J., Ritvo, P., Black, S. E., Freedman, M., Silberfeld, M., & Krahn, M. (2006). Utility-based Quality of Life Measures in Alzheimer's Disease. *Quality of Life Research*, *15*(4), 631–643. https://doi.org/10.1007/s11136-005-4364-8

Oakley, A., Strange, V., Bonell, C., Allen, E., & Stephenson, J. (2006). Process evaluation in randomised controlled trials of complex interventions. *BMJ*, *332*(7538), 413–416. https://doi.org/10.1136/bmj.332.7538.413

Office for National Statistics. (2019). *Deaths registered in England and Wales: 2018*. Office for National Statistics. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsregistrationsummarytables/2018

Orgeta, V., Leung, P., Yates, L., Kang, S., Hoare, Z., Henderson, C., Whitaker, C., Burns, A., Knapp, M., Leroi, I., Moniz-Cook, E. D., Pearson, S., Simpson, S., Spector, A., Roberts, S., Russell, I. T., de Waal, H., Woods, R. T., & Orrell, M. (2015). Individual cognitive stimulation therapy for dementia: A clinical effectiveness and cost-effectiveness pragmatic, multicentre, randomised controlled trial. *Health Technology Assessment*, *19*(64), 1–108. https://doi.org/10.3310/hta19640

Pandis, N. (2011). Sources of bias in clinical trials. *American Journal of Orthodontics and Dentofacial Orthopedics*, *140*(4), 595–596. https://doi.org/10.1016/j.ajodo.2011.06.013

Pannucci, C. J., & Wilkins, E. G. (2010). Identifying and Avoiding Bias in Research: *Plastic and Reconstructive Surgery*, *126*(2), 619–625. https://doi.org/10.1097/PRS.0b013e3181de24bc

Pawson, R. (2013). *The Science of Evaluation: A Realist Manifesto*. SAGE Publications.

Perkins, D. O., Wyatt, R. J., & Bartko, J. J. (2000). Penny-wise and pound-foolish: The impact of measurement error on sample size requirements in clinical trials. *Biological Psychiatry*, *47*(8), 762–766. https://doi.org/10.1016/S0006-3223(00)00837-4

Phillips, S. P., & Hamberg, K. (2016). Doubly blind: A systematic review of gender in randomised controlled trials. *Global Health Action*, *9*(1), 29597. https://doi.org/10.3402/gha.v9.29597

Pitts, M. J., & Miller-Day, M. (2007). Upward turning points and positive rapport-development across time in researcher—Participant relationships. *Qualitative Research*, *7*(2), 177–201. https://doi.org/10.1177/1468794107071409

Pollner, M. (1998). The Effects of Interviewer Gender in Mental Health Interviews. *The Journal of Nervous and Mental Disease*, *186*(6), 369–373.

Portney, L.G. and Watkins, M.P. (2000) *Foundations of clinical research: Applications to practice*. 2nd Edition, Prentice Hall Health, Upper Saddle River, New Jersey.

Post, M. W. (2016). What to Do With "Moderate" Reliability and Validity Coefficients? *Archives of Physical Medicine and Rehabilitation*, *97*(7), 1051–1052. https://doi.org/10.1016/j.apmr.2016.04.001

Prince, M. J. (2016). *World Alzheimer Report 2016 - Improving healthcare for people living with dementia: Coverage, quality and costs now and in the future*. https://www.alz.co.uk/research/world-report-2016

Prince M, Knapp M, Guerchet M, McCrone P, Prina M, Comas-Herrera A, et al. (2014) *Dementia UK: Update*. London: Alzheimer's Society; 2014.

Public Health England. (2018). *Process evaluation*. GOV.UK.

https://www.gov.uk/government/publications/evaluation-in-health-and-well-being-overview/process-evaluation

Quality Of Life. (n.d.). In *Lexico Dictionaries | English*. Retrieved 30 March 2020, from https://www.lexico.com/definition/quality_of_life

Quinn, C., Clare, L., & Woods, B. (2009). The impact of the quality of relationship on the experiences and wellbeing of caregivers of people with dementia: A systematic review. *Aging & Mental Health*, *13*(2), 143–154. https://doi.org/10.1080/13607860802459799

Rabins, P. V., Kasper, J. D., Kleinman, L., Black, B. S., & Patrick, D. L. (1999). Concepts and methods in the development of the ADRQL: An instrument for assessing health-related quality of life in persons with Alzheimer's disease. *Journal of Mental Health and Aging*, *5*(1), 33–48.

Ready, R. E., & Ott, B. R. (2003). Quality of Life measures for dementia. *Health and Quality of Life Outcomes*, 9.

Ready, R. E., Ott, B. R., & Grace, J. (2004). Patient versus informant perspectives of Quality of Life in Mild Cognitive Impairment and Alzheimer's disease. *International Journal of Geriatric Psychiatry*, *19*(3), 256–265. https://doi.org/10.1002/gps.1075

*Risk factors for dementia. Factsheet 450LP.* (p. 17). (2016). Alzheimer's Society. https://www.alzheimers.org.uk/sites/default/files/pdf/factsheet_risk_factors_for_dementia.pdf

Römhild, J., Fleischer, S., Meyer, G., Stephan, A., Zwakhalen, S., Leino-Kilpi, H., Zabalegui, A., Saks, K., Soto-Martin, M., Sutcliffe, C., Rahm Hallberg, I., & Berg, A. (2018). Inter-rater agreement of the Quality of Life-Alzheimer's Disease (QoL-AD) self-rating and proxy rating scale: Secondary analysis of RightTimePlaceCare data. *Health and Quality of Life Outcomes*, *16*(1), 131. https://doi.org/10.1186/s12955-018-0959-y

Rosenkoetter, U., & Tate, R. L. (2018). Assessing Features of Psychometric Assessment Instruments: A Comparison of the COSMIN Checklist with Other Critical Appraisal Tools. *Brain Impairment*, *19*(1), 103–118. https://doi.org/10.1017/BrImp.2017.29

Roter, D. L., & Hall, J. A. (2004). Physician Gender and Patient-Centered Communication: A Critical Review of Empirical Research. *Annual Review of Public Health*, *25*(1), 497–519. https://doi.org/10.1146/annurev.publhealth.25.101802.123134

Rothman, M. L., Beltran, P., Cappelleri, J. C., Lipscomb, J., & Teschendorf, B. (2007). Patient-Reported Outcomes: Conceptual Issues. *Value in Health*, *10*, S66–S75. https://doi.org/10.1111/j.1524-4733.2007.00269.x

Rothstein J (1985) *Measurement and clinical practice: theory and application*. Churchill Livingstone, New York

Rothwell, P. M. (2005). External validity of randomised controlled trials: "To whom do the results of this trial apply?" *The Lancet*, *365*(9453), 82–93. https://doi.org/10.1016/S0140-6736(04)17670-8

Russell, D., Hoare, Z. S. J., Whitaker, Rh., Whitaker, C. J., & Russell, I. T. (2011). Generalized method for adaptive randomization in clinical trials. *Statistics in Medicine*, n/a-n/a. https://doi.org/10.1002/sim.4175

Sands, L. P., Ferreira, P., Stewart, A. L., Brod, M., & Yaffe, K. (2004). What Explains Differences Between Dementia Patients' and Their Caregivers' Ratings of Patients' Quality of Life? *The American Journal of Geriatric Psychiatry; Washington*, *12*(3), 272–280.

Schulz, R., & Martire, L. M. (2004). Family Caregiving of Persons With Dementia: Prevalence, Health Effects, and Support Strategies. *The American Journal of Geriatric Psychiatry; Washington*, *12*(3), 240–249.

Selai, C. E., Trimble, M. R., Rossor, M. N., & Harvey, R. J. (2001). Assessing quality of life in dementia: Preliminary psychometric testing of the Quality of Life Assessment Schedule (QOLAS). *Neuropsychological Rehabilitation*, *11*(3–4), 219–243. https://doi.org/10.1080/09602010042000033

Selai, C., & Trimble, M. R. (1999). Assessing quality of life in dementia. *Aging & Mental Health*, *3*(2), 101–111.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420. https://doi.org/10.1037/0033-2909.86.2.420

Siddiqi, N. (2011). Publication Bias in Epidemiological Studies. *Central European Journal of Public Health*, *19*(2), 118–120. https://doi.org/10.21101/cejph.a3581

Siepmann, T., Spieth, P. M., Kubasch, A. S., Penzlin, A. I., Illigens, B. M.-W., & Barlinn, K. (2016). Randomized controlled trials: A matter of design. *Neuropsychiatric Disease and Treatment*, 1341. https://doi.org/10.2147/NDT.S101938

Simundic, A.-M. (2013). Bias in research. *Biochemia Medica*, 12–15. https://doi.org/10.11613/BM.2013.003

Smith, J., & Noble, H. (2014). Bias in research. *Evidence-Based Nursing*, *17*(4), 100–101. https://doi.org/10.1136/eb-2014-101946

Smith, S., Lamping, D., Banerjee, S., Harwood, R., Foley, B., Smith, P., Cook, J., Murray, J., Prince, M., Levin, E., Mann, A., & Knapp, M. (2005). Measurement of health-related quality of life for people with dementia: Development of a new instrument (DEMQOL) and an evaluation of current methodology. *Health Technology Assessment*, *9*(10). https://doi.org/10.3310/hta9100

Soobiah, C., Tadrous, M., Knowles, S., Blondal, E., Ashoor, H. M., Ghassemi, M., Khan, P. A., Ho, J., Tricco, A. C., & Straus, S. E. (2019). Variability in the validity and reliability of outcome measures identified in a systematic review to assess treatment efficacy of cognitive enhancers for Alzheimer's Dementia. *PLOS ONE*, *14*(4), e0215225. https://doi.org/10.1371/journal.pone.0215225

Spector, A., Orrell, M., Charlesworth, G., & Marston, L. (2016). Factors influencing the person–carer relationship in people with anxiety and dementia. *Aging & Mental Health*, *20*(10), 1055–1062. https://doi.org/10.1080/13607863.2015.1063104

Spruytte, N., Audenhove, C., Lammertyn, F., & Storms, G. (2002). The quality of the caregiving relationship in informal care for older adults with dementia and chronic psychiatric patients.

*Psychology and Psychotherapy: Theory, Research and Practice*, *75*(3), 295–311. https://doi.org/10.1348/147608302320365208

Stahl, G. (2016). Relationship-Building in Research: Gendered Identity Construction in Researcher-Participant Interaction. In M. R. M. Ward (Ed.), *Studies in Qualitative Methodology* (Vol. 14, pp. 145–165). Emerald Group Publishing Limited. https://doi.org/10.1108/S1042-319220160000014020

Steeman, E., Godderis, J., Grypdonck, M., De Bal, N., & De Casterlé, B. D. (2007). Living with dementia from the perspective of older people: Is it a positive story? *Aging & Mental Health*, *11*(2), 119–130. https://doi.org/10.1080/13607860600963364

Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), Handbook of experimental psychology (p. 1–49). Wiley.

Streiner, D. L. (2008). *Health Measurement Scales*. OUP Oxford.

Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012

The BMJ. (n.d.). *Chapter 4. Measurement error and bias*. Retrieved 29 March 2020, from https://www.bmj.com/about-bmj/resources-readers/publications/epidemiology-uninitiated/4-measurement-error-and-bias

The Good Care Group. (n.d.). *Why care continuity is crucial for people with Dementia*. The Good Care Group. Retrieved 30 March 2020, from https://www.thegoodcaregroup.com/news/why-care-continuity-crucial-people-dementia/

Thomas, G. P. A., Saunders, C. L., Roland, M. O., & Paddison, C. A. M. (2015). Informal carers' health-related quality of life and patient experience in primary care: Evidence from 195,364 carers in England responding to a national survey. *BMC Family Practice*, *16*(1), 62. https://doi.org/10.1186/s12875-015-0277-y

Thompson, L., & Kingston, P. (2004). Measures to assess the quality of life for people with advanced dementia: Issues in measurement and conceptualisation. *Quality in Ageing and Older Adults*, *5*(4), 29–39. https://doi.org/10.1108/14717794200400022

Thurnell-Read, T. (2016). Masculinity, Age and Rapport in Qualitative Research. In M. R. M. Ward (Ed.), *Studies in Qualitative Methodology* (Vol. 14, pp. 23–41). Emerald Group Publishing Limited. https://doi.org/10.1108/S1042-319220160000014014

*Treatments*. (n.d.). Alzheimer's Society. Retrieved 29 March 2020, from https://www.alzheimers.org.uk/about-dementia/treatments

Treweek, S., Bevan, S., Bower, P., Campbell, M., Christie, J., Clarke, M., Collett, C., Cotton, S., Devane, D., El Feky, A., Flemyng, E., Galvin, S., Gardner, H., Gillies, K., Jansen, J., Littleford, R., Parker, A., Ramsay, C., Restrup, L., … Williamson, P. R. (2018). Trial Forge Guidance 1: What is a Study Within A Trial (SWAT)? *Trials*, *19*(1), 139. https://doi.org/10.1186/s13063-018-2535-5

Trial Forge. (n.d.). *Trial Forge—A systematic way to improve trial efficiency.* Trial Forge. Retrieved 29 March 2020, from https://www.trialforge.org/

Trigg, R., Skevington, S. M., & Jones, R. W. (2007). How Can We Best Assess the Quality of Life of People With Dementia? The Bath Assessment of Subjective Quality of Life in Dementia (BASQID). *The Gerontologist*, *47*(6), 789–797. https://doi.org/10.1093/geront/47.6.789

Trigg, Richard, Watts, S., Jones, R., & Tod, A. (2011). Predictors of quality of life ratings from persons with dementia: The role of insight. *International Journal of Geriatric Psychiatry*, *26*(1), 83–91. https://doi.org/10.1002/gps.2494

Tripepi, G., Jager, K. J., Dekker, F. W., Wanner, C., & Zoccali, C. (2008). Bias in clinical research. *Kidney International*, *73*(2), 148–153. https://doi.org/10.1038/sj.ki.5002648

Turner, R. R., Quittner, A. L., Parasuraman, B. M., Kallich, J. D., & Cleeland, C. S. (2007). Patient-Reported Outcomes: Instrument Development and Selection Issues. *Value in Health*, *10*, S86–S93. https://doi.org/10.1111/j.1524-4733.2007.00271.x

*Types of dementia*. (n.d.). Alzheimer's Society. Retrieved 29 March 2020, from

https://www.alzheimers.org.uk/about-dementia/types-dementia

Walker, A., & Lowenstein, A. (2009). European perspectives on quality of life in old age. *European Journal of Ageing*, *6*(2), 61–66. https://doi.org/10.1007/s10433-009-0117-9

Walker, I. (2008). *Null hypothesis testing and effect sizes*. Statistics for Psychology. https://people.bath.ac.uk/pssiw/stats2/page2/page14/page14.html

Weiner, M., Martin-Cook, K., Svetlik, D., Saine, K., Foster, B., & Fontaine, C. (1999). The quality of life in late-stage dementia (QUALID) scale. *Journal of the American Medical Directors Association*, *1*, 114–116. https://doi.org/10.1037/t00432-000

West, B. T., & Blom, A. G. (2016). Explaining Interviewer Effects: A Research Synthesis. *Journal of Survey Statistics and Methodology*, smw024. https://doi.org/10.1093/jssam/smw024

Whitaker, R., Fossey, J., Ballard, C., Orrell, M., Moniz-Cook, E., Woods, R. T., Murray, J., Stafford, J., Knapp, M., Romeo, R., Carlton, B., Testad, I., & Khan, Z. (2014). Improving Well-being and Health for People with Dementia (WHELD): Study protocol for a randomised controlled trial. *Trials*, *15*(1), 284. https://doi.org/10.1186/1745-6215-15-284

Williams, C. L., & Heikes, E. J. (1993). The Importance of Researcher's Gender in the In-Depth Interview: Evidence from Two Case Studies of Male Nurses. *Gender and Society*, *7*(2), 280–291. JSTOR.

Wittenberg, R., Hu, B., Barraza-Araiza, L., & Rehill, A. (2019). *Projections of older people with dementia and costs of dementia care in the United Kingdom, 2019–2040*. Care Policy and Evaluation Centre, London School of Economics and Political Science.

Women and Dementia: A Marginalised Majority. (n.d.). *Alzheimer's Research UK*. Retrieved 29 March 2020, from https://www.alzheimersresearchuk.org/about-us/our-influence/policy-work/reports/women-dementia/

Woods, B., Thorgrimsen, L., Spector, A., Royan, L., & Orrell, M. (2006). Improved quality of life and cognitive stimulation therapy in dementia. *Aging & Mental Health*, *10*(3), 219–226. https://doi.org/10.1080/13607860500431652

Woods, R., Bruce, E., Edwards, R., Elvish, R., Hoare, Z., Hounsome, B., Keady, J., Moniz-Cook, E., Orgeta, V., Orrell, M., Rees, J., & Russell, I. (2012). REMCARE: Reminiscence groups for people with dementia and their family caregivers – effectiveness and cost-effectiveness pragmatic multicentre randomised trial. *Health Technology Assessment*, *16*(48). https://doi.org/10.3310/hta16480

Woods, R. T., Nelis, S. M., Martyr, A., Roberts, J., Whitaker, C. J., Markova, I., Roth, I., Morris, R., & Clare, L. (2014). What contributes to a good quality of life in early dementia? Awareness and the QoL-AD: a cross-sectional study. *Health and Quality of Life Outcomes*, *12*(1), 94. https://doi.org/10.1186/1477-7525-12-94

World Health Organisation. (n.d.). *WHOQOL: Measuring Quality of Life*. WHO. Retrieved 29 March 2020, from https://www.who.int/healthinfo/survey/whoqol-qualityoflife/en/

World Health Organisation. (2017). *Global action plan on the public health response to dementia 2017—2025* (p. 52). WHO. https://apps.who.int/iris/bitstream/handle/10665/259615/9789241513487-eng.pdf;jsessionid=23B00EFDE571F9E3FFDC2DAE7D7A9031?sequence=1

Wright, J., Foster, A., Cooper, C., Sprange, K., Walters, S., Berry, K., Moniz-Cook, E., Loban, A., Young, T. A., Craig, C., Dening, T., Lee, E., Beresford-Dent, J., Thompson, B. J., Young, E., Thomas, B. D., & Mountain, G. (2019). Study protocol for a randomised controlled trial assessing the clinical and cost-effectiveness of the Journeying through Dementia (JtD) intervention compared to usual care. *BMJ Open*, *9*(9), e029207. https://doi.org/10.1136/bmjopen-2019-029207

Wu, Y.-T., Fratiglioni, L., Matthews, F. E., Lobo, A., Breteler, M. M. B., Skoog, I., & Brayne, C. (2016). Dementia in western Europe: Epidemiological evidence and implications for policy making. *The Lancet Neurology*, *15*(1), 116–124. https://doi.org/10.1016/S1474-4422(15)00092-7

Yang, F., Dawes, P., Leroi, I., & Gannon, B. (2018). Measurement tools of resource use and quality of life in clinical trials for dementia or cognitive impairment interventions: A systematically

conducted narrative review. *International Journal of Geriatric Psychiatry*, *33*(2), e166–e176. https://doi.org/10.1002/gps.4771

Zwaanswijk, M., Peeters, J. M., van Beek, A. P., Meerveld, J. H., & Francke, A. L. (2013). Informal Caregivers of People with Dementia: Problems, Needs and Support in the Initial Stage and in Subsequent Stages of Dementia: A Questionnaire Survey. *The Open Nursing Journal*, *7*, 6–13. https://doi.org/10.2174/1874434601307010006

# Appendices

Appendix 1: Common types of biases in randomised controlled trials.

| **Type of bias** | **Definition of bias** | **Implication** |
|---|---|---|
| Publication bias | Some research is published and some is not published where it is factors other than the quality of the study that influences the decision to publish or not such as the outcome of the research, direction of the results or the topic of the study. | Only certain studies of interventions are accessible which impacts conclusions of meta-analysis and systematic reviews of these interventions |
| Analysis bias | Through data manipulation or inappropriate analysis methods adopted. | Inaccurate estimate of treatment effect which could lead to incorrect conclusions of interventions or treatments. |
| Contamination bias | Participants in one group inadvertently receive effects from the other group through association. This type of bias tends to occur in certain circumstances where patients allocated to experimental or control mix with one another. | Inaccurate estimate of treatment effect which could lead to incorrect conclusions of interventions or treatments. |
| Compliance bias | Participants do not adhere to the study intervention, for example by not taking study medication or attending therapy sessions. | Inaccurate estimate of treatment effect which could lead to incorrect conclusions of interventions or treatments. |
| Attrition bias | Relates to the way in which subject withdrawals are handled for analysis and is a particular issue if withdrawals differ systematically between comparison arms | Inaccurate estimate of treatment effect which could lead to incorrect conclusions of interventions or treatments. |
| Selection bias | Method of selecting participants for the study results in either the baseline characteristics having systematic differences between comparison groups or a non-representative study sample of the target population is collected. | Results of the study are ungeneralizable. |
| Performance bias | Occurs if the intervention delivery, follow up care or other contributing factors differ for certain patients and is a particular issue when overall patients in one group receive a higher standard of treatment than those in the other group). | Inaccurate estimate of treatment effect which could lead to incorrect conclusions of interventions or treatments. |
| Detection bias | When researchers collecting follow up data know the treatment allocation of the participant and the data collected is affected by this knowledge which tends to be a particular problem if the outcomes involve subjective evaluations. | Inaccurate estimate of treatment effect which could lead to incorrect conclusions of interventions or treatments. |
| Reporting bias | Occurs from the way in which authors report their study findings in a less honest way that supports a certain view point. | Summaries of results of the study publications may be inaccurate and conclusions drawn of the intervention effect may be distorted. |

**Appendix 2: Table of outcome measure details**

| Outcome measure | Definition | Scoring | | | | | Subscales | Missing value rules |
|---|---|---|---|---|---|---|---|---|
| | | Number of items | Items scored on | Scoring calculation | Final score range | Score direction | | |
| QCPR | Assess caregiving relationship. | 14 items | 5-point Likert-scale. 1-5 ('Totally not agree' - 'Totally agree')<br><br>Conflict subscale items scored in reverse. | Final score calculated by summing items. | 14 to 70 | Higher score indicates better perceived relationship | <u>Warmth</u><br>Items; 1, 4, 5, 6, 7, 9, 12, 14<br><br><u>Conflict & criticism</u><br>Items; 2,3,8,10,11,13 | None. |
| QoL-AD | Assess quality of life. | 13 items | 4-point Likert-scale<br><br>1–4 ('poor' - 'Excellent') | Final score calculated by summing individual components. | 13 to 52 | Higher score on scale indicates a better quality of life | None. | Up to two missing items are replaced by the remaining mean scores. |

Appendix 3: Completion rates for raw scores of outcome measures.

| Number of missing items | Researcher Attendance Follow-up 1 | | | | | | Researcher Attendance Follow-up 2 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Baseline | | | Follow-up 1 | | | Baseline | | | | Follow-up 2 | | | |
| | Total | Same | Different | Total | Same | Different | Total | Same | Two Different | Three Different | Total | Same | Two Different | Three Different |
| **PwD QCPR** | | | | | | | | | | | | | | |
| **None** | **308** | 149 | 159 | **293** | 138 | 155 | **308** | 108 | 124 | 76 | **288** | 98 | 114 | 76 |
| **1 - 13** | **18** | 9 | 9 | **27** | 16 | 10 | **18** | 9 | 3 | 6 | **22** | 10 | 9 | 3 |
| **All (14)** | **4** | 2 | 2 | **10** | 6 | 4 | **4** | 1 | 2 | 1 | **20** | 10 | 6 | 1 |
| **PwD QoL-AD** | | | | | | | | | | | | | | |
| **None** | **274** | 127 | 147 | **250** | 113 | 137 | **274** | 92 | 223 | 70 | **238** | 79 | 95 | 64 |
| ***1 - 2** | **43** | 23 | 20 | **59** | 32 | 27 | **43** | 17 | 25 | 12 | **54** | 20 | 23 | 11 |
| **3 - 12** | **11** | 9 | 2 | **15** | 11 | 4 | **11** | 9 | 3 | 1 | **20** | 11 | 5 | 4 |
| **All (13)** | **2** | 1 | 1 | **6** | 4 | 2 | **2** | 1 | 1 | 0 | **18** | 8 | 6 | 4 |
| **Carer QCPR** | | | | | | | | | | | | | | |
| **None** | **321** | 156 | 165 | **312** | 149 | 163 | **321** | 115 | 128 | 78 | **321** | 117 | 126 | 78 |
| **1 - 13** | **9** | 5 | 4 | **15** | 11 | 4 | **9** | 3 | 1 | 5 | **9** | 1 | 3 | 5 |
| **All (14)** | **0** | 0 | 0 | **3** | 0 | 3 | **0** | 0 | 0 | 0 | **0** | 0 | 0 | 0 |
| **0=Carer proxy QoL-AD** | | | | | | | | | | | | | | |
| **None** | **308** | 149 | 159 | **304** | 147 | 157 | **308** | 108 | 127 | 73 | **300** | 108 | 117 | 75 |
| ***1 - 2** | **18** | 8 | 10 | **19** | 10 | 9 | **18** | 7 | 2 | 9 | **29** | 10 | 12 | 7 |
| **3 - 12** | **4** | 3 | 1 | **4** | 1 | 3 | **4** | 3 | 0 | 1 | **1** | 0 | 0 | 1 |
| **All (13)** | **0** | 0 | 0 | **3** | 2 | 1 | **0** | 0 | 0 | 0 | **0** | 0 | 0 | 0 |

*Indicates that the total score will still be calculated for those participants according to the measures missing data rule.

Appendix 4: List of assumption checks and methods to evaluate for T-tests, ANOVAs Pearson's correlations and Pearson's association

| Assumption | Check |
|---|---|
| **Independent samples t-test and ANOVA** | |
| 1. You have one dependent variable that is measured at the continuous level | Study design - If assumption doesn't hold a different test is required. |
| 2. You have one independent variable that consists of two categorical, independent groups | Study design - If assumption doesn't hold a different test is required. |
| 3. You should have independence of observations | Study design - If assumption doesn't hold a different test is required. |
| 4. There should be no significant outliers in the two groups of your independent variable in terms of the dependent variable | Box plots for the independent variable on the outcome measures at each level of the independent factor |
| 5. Your dependent variable should be approximately normally distributed for each group of the independent variable | Normality Q-Q plots of the dependent variable at each level of the independent variable |
| 6. You have homogeneity of variances (i.e., the variance is equal in each group of your independent variable) | P is greater than 0.05 assumption holds and the assumed equal variances results are reported.<br><br>P is less than 0.05 the assumption is violated and the 'assumed unequal variance' results reported for t-tests and the Welch test results for ANOVAs. |
| **Pearson's Correlation** | |
| 1. Your two variables should be measured on a continuous scale | Study design - If assumption doesn't hold a different test is required. |
| 2. Your two continuous variables should be paired, which means that each case has two values: one for each variable. | Study design - If assumption doesn't hold a different test is required. |
| 3. There needs to be a **linear relationship** between the two variables. | Scatterplot of the two variables |
| 4. There should be no significant outliers. | Scatterplot of the two variables |
| 5. There should be bivariate normality | Normality Q-Q plots of the variables |
| **Chai square Test of independence** | |
| 1. You have two nominal variables. | Study design - If assumption doesn't hold a different test is required. |
| 2. You should have independence of observations. | Study design - If assumption doesn't hold a different test is required. |
| 3. You must have data obtained with cross-sectional sampling | Study design - If assumption doesn't hold a different test is required. |
| 4. All cells should have expected counts greater than or equal to five. | Cross tabulation table |

## Appendix 5: Results from T-tests, ANOVAs, Pearson's correlation and Pearson's Chi-square tests

**Table 1:** T-test and ANOVA of the researcher attendance and researcher gender attendance

**Researcher Attendance at Follow-up 1**

| Outcome Measure | One Researcher | | Two Researchers | | Mean Difference (95% CI) | T- test (df), Significance |
|---|---|---|---|---|---|---|
| | N | Mean (SD) | N | Mean (SD) | | |
| **PwD QCPR** | 138 | 58.2 (6.82) | 155 | 58.1 (6.45) | 0.17 (-1.35, 1.70) | t(291) = 0.22 p = 0.82 |
| **PwD QoL-AD** | 145 | 36.8 (6.11) | 164 | 37.2 (5.36) | -0.39 (-1.68, 0.89) | t(307) = -0.61 p = 0.55 |
| **Carer QCPR** | 149 | 53.0 (9.29) | 163 | 54.2 (8.37) | -1.22 (-3.19, 0.75) | t(310) = -1.22 p = 0.22 |
| **Carer proxy QoL-AD** | 157 | 30.9 (6.31) | 166 | 31.2 (6.18) | -0.30 (-1.67, 1.07) | t(321) = -0.43 p = 0.67 |

**Researcher Attendance at Follow-up 2**

| Outcome Measure | One Researcher | | Two Researchers | | Three Researchers | | F- test (df), Significance |
|---|---|---|---|---|---|---|---|
| | N | Mean (SD) | N | Mean (SD) | N | Mean (SD) | |
| **PwD QCPR** | 98 | 57.4 (6.41) | 114 | 56.7 (6.09) | 76 | 58.2 (7.05) | F(2,285) = 1.309 p = 0.27 |
| **PwD QoL-AD** | 99 | 35.5 (5.48) | 118 | 37.2 (5.05) | 75 | 36.5 (6.03) | F(2,289) = 2.403 p = 0.09 |
| **Carer QCPR** | 117 | 52.2 (9.97) | 126 | 53.0 (9.47) | 78 | 54.6 (8.98) | F(2,318) = 1.479 p = 0.23 |
| **Carer proxy QoL-AD** | 118 | 29.7 (5.90) | 129 | 30.7 (6.40) | 82 | 30.1 (5.72) | F(2,326) = 0.875 p = 0.42 |

**Researcher Gender Attendance at Follow-up 1**

| Outcome Measure | Same Researcher (same gender) | | Different Researcher Same gender | | Different Researchers Different genders | | F- test (df), Significance |
|---|---|---|---|---|---|---|---|
| | N | Mean (SD) | N | Mean (SD) | N | Mean (SD) | |
| **PwD QCPR** | 138 | 58.2 (6.82) | 107 | 59.0 (6.33) | 44 | 56.0 (6.22) | F(2,286) = 3.28 *p = 0.04 |
| **PwD QoL-AD** | 145 | 36.8 (6.11) | 114 | 37.0 (5.75) | 46 | 37.6 (4.40) | F(2, 302) = 0.33 p = 0.72 |
| **Carer QCPR** | 149 | 53.0 (9.29) | 112 | 53.3 (8.73) | 47 | 56.5 (7.30) | F(2, 305) = 2.96 *p = 0.05 |
| **Carer proxy QoL-AD** | 157 | 30.9 (6.31) | 115 | 30.3 (5.98) | 47 | 33.5 (6.28) | F(2, 316) = 4.40 **p = 0.01 |

**Researcher Gender Attendance at Follow-up 2**

| Outcome Measure | Same Researcher (same gender) | | Two different researchers Same gender | | All three different researchers Same gender | | Different researchers (2 or 3) Different genders | | F- test (df), Significance |
|---|---|---|---|---|---|---|---|---|---|
| | N | Mean (SD) | N | Mean (SD) | N | Mean (SD) | N | Mean (SD) | |
| **PwD QCPR** | 98 | 57.4 (6.41) | 72 | 57.3 (5.91) | 52 | 58.3 (5.74) | 63 | 56.4 (7.78) | F(3, 281) = 0.78 p = 0.51 |
| **PwD QoL-AD** | 99 | 35.5 (5.48) | 73 | 37.7 (5.11) | 51 | 35.9 (5.90) | 66 | 36.8 (5.41) | F(3, 285) = 2.58 *p = 0.05 |
| **Carer QCPR** | 117 | 52.1 (9.97) | 78 | 52.8 (9.13) | 55 | 54.2 (8.81) | 67 | 53.8 (9.96) | F(3, 313) = 0.75 p = 0.52 |
| **Carer proxy QoL-AD** | 118 | 29.7 (5.90) | 81 | 30.0 (6.15) | 57 | 30.1 (5.56) | 69 | 31.1 (6.23) | F(3, 321) = 0.82 p = 0.49 |

**Significant at the 0.05 level. **Significant at the 0.01 level.*

**Table 2:** T-test results on Gender, Marital Status and allocation for the participant and carer data

| | Follow-up 1 | | | | | | Follow-up 2 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Gender** | **Female** | | **Male** | | **Mean Difference (95% CI)** | **T- test (df), Significance** | **Female** | | **Male** | | **Mean Difference (95% CI)** | **T- test (df), Significance** |
| | **N** | **Mean (SD)** | **N** | **Mean (SD)** | | | **N** | **Mean (SD)** | **N** | **Mean (SD)** | | |
| **PwD QCPR** | 141 | 58.1 (6.56) | 151 | 58.1 (6.71) | 0.01 (-1.52, 1.54) | t(209) = 0.01 p = 0.99 | 138 | 57.2 (6.52) | 149 | 57.5 (6.45) | -0.25 (-1.75, 1.26) | t(285) = -0.32 p = 0.75 |
| **PwD QoL-AD** | 148 | 37.8 (5.43) | 160 | 36.4 (5.92) | 1.315 (0.04, 2.59) | t(306) = 2.03 p = *0.04 | 138 | 36.9 (5.37) | 153 | 36.0 (5.57) | 0.850 (-0.42, 2.12) | t(289) = 1.32 p = 0.19 |
| **Carer QCPR** | 208 | 52.2 (8.85) | 103 | 56.5 (8.11) | -4.33 (-6.37, -2.28) | t(309) = -4.17 p = **<0.01 | 214 | 51.9 (9.63) | 106 | 55.6 (8.88) | -3.65 (-5.85, -1.45) | t(318) = -3.27 p = **<0.01 |
| **Carer proxy QoL-AD** | 215 | 30.3 (6.35) | 107 | 32.6 (5.74) | -2.28 (-3.71, -0.85) | t(320) = -3.13 p = **<0.01 | 221 | 29.5 (5.93) | 107 | 31.7 (6.5) | -2.28 (-3.66, -0.89) | t(326) = -3.24 p = **<0.01 |
| **PwD gender on carer QCPR** | 156 | 55.5 (7.92) | 155 | 51.7 (9.32) | 3.76 (1.83, 5.69) | t(301) = 3.83 p = **<0.01 | 158 | 54.5 (9.06) | 162 | 51.8 (9.82) | 2.74 (0.66, 4.81) | t(318) = 2.59 **p = 0.01 |
| **PwD gender on carer proxy QoL-AD** | 159 | 31.5 (5.99) | 163 | 30.7 (6.47) | 0.79 (-0.57, 2.16) | t(320) = 1.14 p = 0.26 | 161 | 30.4 (6.13) | 167 | 30.0 (5.99) | 0.48 (-0.86, 1.77) | t(326) = 0.68 p = 0.50 |
| | **Follow-up 1** | | | | | | **Follow-up 2** | | | | | |
| **Marital Status** | **Spousal** | | **Non-Spousal** | | **Mean Difference (95% CI)** | **T- test (df), Significance** | **Spousal** | | **Non-Spousal** | | **Mean Difference (95% CI)** | **T- test (df), Significance** |
| | **N** | **Mean (SD)** | **N** | **Mean (SD)** | | | **N** | **Mean (SD)** | **N** | **Mean (SD)** | | |
| **PwD QCPR** | 219 | 58.3 (6.79) | 72 | 57.7 (6.18) | 0.63 (-1.15, 2.40) | t(289) = 0.69 p = 0.50 | 214 | 57.4 (6.59) | 73 | 57.2 (6.18) | 0.24 (-1.49, 1.97) | t(285) = 0.28 p = 0.78 |
| **PwD QoL-AD** | 228 | 37.2 (5.75) | 79 | 36.9 (5.59) | 0.31 (-1.16, 1.77) | t(305) = 0.41 p = 0.68 | 218 | 36.6 (5.62) | 72 | 36.0 (5.0) | 0.65 (-0.81, 2.11) | t(288) = 0.88 p = 0.38 |
| **Carer QCPR** | 274 | 53.7 (8.88) | 35 | 53.4 (8.36) | 0.30 (-2.81, 3.42) | t(307) = 0.19 p = 0.85 | 282 | 53.0 (9.49) | 36 | 53.9 (10.48) | -0.97 (-4.31, 2.37) | t(316)= -0.57 p = 0.57 |
| **Carer proxy QoL-AD** | 283 | 31.0 (6.01) | 36 | 32.4 (7.83) | -1.40 (-3.57, 0.77) | t(317) = -1.27 p = 0.21 | 289 | 30.1 (5.83) | 36 | 31.0 (7.83) | -0.82 (-2.94, 1.29) | t(323) = -0.77 p = 0.44 |
| | **Follow-up 1** | | | | | | **Follow-up 2** | | | | | |
| **Allocation** | **RYCT** | | **TAU** | | **Mean Difference (95% CI)** | **T- test (df), Significance** | **RYCT** | | **TAU** | | **Mean Difference (95% CI)** | **T- test (df), Significance** |
| | **N** | **Mean (SD)** | **N** | **Mean (SD)** | | | **N** | **Mean (SD)** | **N** | **Mean (SD)** | | |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **PwD QCPR** | 172 | 58.4 (6.45) | 121 | 57.7 (6.85) | 0.75 (-0.8, 2.29) | t(291) = 0.95 p = 0.34 | 166 | 57.5 (6.28) | 122 | 57.0 (6.74) | 0.47 (-1.05, 1.99) | t(286) = 0.61 p = 0.54 |
| **PwD QoL-AD** | 183 | 37.1 (5.75) | 126 | 36.9 (5.69) | 0.20 (-1.10, 1.51) | t(307) = 0.31 p = 0.76 | 171 | 36.8 (5.58) | 121 | 35.9 (5.32) | 0.85 (-0.43, 2.13) | t(290) = 1.31 p = 0.19 |
| **Carer QCPR** | 182 | 52.9 (9.05) | 130 | 54.5 (8.46) | -1.60 (-3.59, 0.39) | T(310) = -1.57 P = 0.11 | 185 | 53.0 (9.78) | 136 | 53.2 (9.28) | -0.15 (-2.28, 1.97) | t(319) = -0.14 p = 0.89 |
| **Carer proxy QoL-AD** | 189 | 30.6 (5.93) | 134 | 31.7 (6.61) | -1.16 (-2.54, 0.22) | T(321) = -1.65 p = 0.10 | 191 | 29.8 (5.86) | 138 | 30.7 (6.31) | -0.89 (-2.22, 0.44) | t(327) = -1.32 p = 0.19 |

**Significant at the 0.05 level. **Significant at the 0.01 level.*

**Table 3:** ANOVA conducted on Wave at follow-up 1 and follow-up 2 on the participant and carer data

| **Wave** | **Wave 1** | | **Wave 2** | | **Wave 3** | | **Wave 4** | | **Wave 5** | | **F- test (df), Significance** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Outcome Measure** | **N** | **Mean (SD)** | **N** | **Mean (SD)** | **N** | **Mean (SD)** | **N** | **Mean (SD)** | **N** | **Mean (SD)** | |
| | **Follow-up 1** | | | | | | | | | | |
| **PwD QCPR** | 94 | 57.0 (5.98) | 70 | 59.0 (6.81) | 78 | 58.7 (6.78) | 46 | 58.7 (7.31) | 5 | 54.2 (2.59) | F(4,288) = 1.67 p = 0.16 |
| **PwD QoL-AD** | 89 | 36.6 (5.21) | 70 | 37.8 (5.39) | 80 | 35.8 (6.08) | 48 | 36.4 (5.36) | 5 | 38.6 (2.70) | F4,287) = 0.51 p = 0.73 |
| **Carer QCPR** | 98 | 53.8 (8.41) | 72 | 53.8 (9.72) | 84 | 53.9 (8.86) | 53 | 52.2 (8.71) | 5 | 55.8 (4.49) | F(4,307) = 0.44 p = 0.78 |
| **Carer proxy QoL-AD** | 101 | 30.7 (5.86) | 77 | 31.0 (6.38) | 86 | 31.4 (6.15) | 54 | 31.1 (7.10) | 5 | 32.0 (4.53) | F(4,318) = 0.18 p = 0.95 |
| | **Follow-up 2** | | | | | | | | | | |
| **PwD QCPR** | 88 | 57.1 (6.02) | 70 | 57.7 (6.82) | 80 | 57.6 (5.99) | 45 | 56.9 (7.88) | 5 | 57.2 (3.83) | F(4,283) = 0.16 p = 0.96 |
| **PwD QoL-AD** | 96 | 37.3 (5.92) | 77 | 37.4 (5.37) | 81 | 36.8 (5.91) | 50 | 36.5 (5.85) | 5 | 36.5 (3.07) | F(4,304) = 0.32 p = 0.83 |
| **Carer QCPR** | 100 | 52.9 (9.19) | 77 | 53.4 (10.37) | 87 | 54.3 (9.17) | 52 | 50.7 (9.70) | 5 | 54.8 (6.83) | F(4,316) = 1.298 p = 0.271 |
| **Carer proxy QoL-AD** | 101 | 30.1 (6.36) | 82 | 29.8 (5.76) | 87 | 30.6 (5.91) | 54 | 29.9 (6.26) | 5 | 33.2 (6.14) | F(4,324) = 0.53 p = 0.72 |

**Significant at the 0.05 level. **Significant at the 0.01 level.*

**Table 4:** ANOVA conducted on Centre at follow-up 1 and follow-up 2 on the participant and carer data

| Centre | Follow-up 1 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Outcome Measure | Bangor | | Bradford | | Gwent | | Hull | | London North | | London South | | Manchester | | F- test (df), Significance |
| | Mean (SD) | N | Mean (SD) | N | Mean (SD) | N | Mean (SD) | N | Mean (SD) | N | Mean (SD) | N | Mean (SD) | N | |
| PwD QCPR | 57.4 (6.15) | 43 | 57.9 (7.81) | 35 | 60.9 (5.60) | 9 | 57.7 (7.26) | 45 | 58.1 (7.12) | 55 | 57.6 (5.95) | 43 | 59.0 (5.92) | 63 | F(6,286) = 0.62 p = 0.71 |
| PwD QoL-AD | 37.0 (5.17) | 47 | 34.3 (5.75) | 27 | 39.0 (2.80) | 14 | 34.8 (5.32) | 45 | 36.0 (5.60) | 50 | 36.2 (6.16) | 46 | 38.0 (5.00) | 63 | F(6,285) = 2.96 p = **0.01 |
| Carer QCPR | 51.8 (8.77) | 56 | 52.8 (10.37) | 36 | 56.3 (8.13) | 13 | 55.0 (8.01) | 45 | 53.1 (10.1) | 56 | 55.5 (7.27) | 45 | 53.2 (8.30) | 61 | F(6,305) = 1.24 p = 0.29 |
| Carer proxy QoL-AD | 31.0 (6.55) | 57 | 30.0 (6.26) | 38 | 32.7 (6.13) | 14 | 32.7 (5.88) | 46 | 31.2 (6.66) | 58 | 30.3 (6.28) | 46 | 30.6 (5.75) | 64 | F(6,316) = 1.056 p = 0.39 |
| **Follow-up 2** | | | | | | | | | | | | | | | |
| Outcome Measure | Bangor | | Bradford | | Gwent | | Hull | | London North | | London South | | Manchester | | F- test (df), Significance |
| | Mean (SD) | N | Mean (SD) | N | Mean (SD) | N | Mean (SD) | N | Mean (SD) | N | Mean (SD) | N | Mean (SD) | N | |
| PwD QCPR | 56.4 (6.17) | 42 | 55.8 (7.65) | 31 | 59.7 (5.93) | 14 | 57.5 (6.85) | 45 | 58.2 (7.33) | 50 | 57.8 (6.02) | 46 | 56.9 (5.31) | 60 | F(6,281) = 0.97 p = 0.44 |
| PwD QoL-AD | 36.4 (6.80) | 52 | 35.8 (5.92) | 32 | 40.5 (2.66) | 13 | 36.4 (5.23) | 43 | 36.5 (5.80) | 57 | 37.6 (5.18) | 48 | 38.0 (5.51) | 64 | F(6,302) = 1.75 p = 0.11 |
| Carer QCPR | 50.9 (9.90) | 57 | 54.3 (11.2) | 36 | 55.8 (7.26) | 14 | 52.7 (9.67) | 46 | 53.4 (9.93) | 57 | 55.5 (7.49) | 45 | 52.2 (9.37) | 66 | F(6, 314) = 1.39 p = 0.22 |
| Carer proxy QoL-AD | 30.2 (6.35) | 57 | 29.6 (5.89) | 39 | 32.1 (6.62) | 14 | 31.5 (5.57) | 46 | 29.4 (6.28) | 58 | 29.9 (5.14) | 49 | 30.0 (6.56) | 66 | F(6,322) = 0.83 p = 0.55 |

**Significant at the 0.05 level. **Significant at the 0.01 level.*

**Table 5:** Pearson's correlation co-efficient for age on the QCPR and QoL-AD at each time point

| | | Follow-up 1 | | | Follow-up 2 | | |
|---|---|---|---|---|---|---|---|
| | | N | Pearson's Correlation | P value (two tailed) | N | Pearson's Correlation | P value (two tailed) |
| Age of PwD | QCPR | 293 | 0.01 | 0.85 | 288 | 0.12* | 0.04 |
| | QoL-AD | 309 | <0.01 | 0.98 | 292 | 0.02 | 0.70 |
| Age of Carer | QCPR | 312 | 0.10 | 0.08 | 321 | 0.06 | 0.26 |
| | Proxy QoL-AD | 323 | 0.16** | <0.01 | 329 | 0.10 | 0.06 |
| Age of PwD | QCPR | 312 | -0.01 | 0.84 | 321 | 0.02 | 0.77 |
| | Proxy QoL-AD | 323 | -0.07 | 0.20 | 329 | -0.14** | 0.01 |
| PwD baseline | QCPR | 279 | 0.56** | <0.01 | 272 | 0.37** | <0.01 |
| | QoL-AD | 300 | 0.66** | <0.01 | 283 | 0.54** | <0.01 |
| Carer baseline | QCPR | 305 | 0.75** | <0.01 | 312 | 0.65** | <0.01 |
| | Proxy QoL-AD | 321 | 0.75** | <0.01 | 325 | 0.68** | <0.01 |

**Correlation is significant at the 1% level. *Correlation is significant at the 5% level.

**Table 7:** Pearson's Chi-square test of independence between researcher variables and categorical factors

| Variables | Follow-up 1 | | | | | Follow-up 2 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | N | Chi-Square Value | DF | P value | Cramer's V | N | Chi-Square Value | DF | P value | Cramer's V |
| **Researcher attendance** | | | | | | | | | | |
| **PwD gender** | 329 | 0.50 | 1 | 0.48 | 0.04 | 329 | 0.57 | 2 | 0.75 | 0.04 |
| **Carer gender** | 329 | 0.01 | 1 | 0.91 | <0.01 | 329 | 0.47 | 2 | 0.79 | 0.04 |
| **PwD Marital Status** | 328 | 0.20 | 1 | 0.66 | 0.03 | 328 | 0.31 | 2 | 0.86 | 0.03 |
| **Carer Marital Status** | 326 | 0.26 | 1 | 0.61 | 0.03 | 326 | 3.46 | 2 | 0.18 | 0.10 |
| **Centre** | 330 | **77.15 | 6 | <0.01 | 0.48 | 330 | **146.17 | 12 | <0.01 | 0.47 |
| **Wave** | 330 | 6.54 | 4 | 0.16 | 0.14 | 330 | **19.41 | 8 | 0.01 | 0.17 |
| **Allocation** | 330 | 0.04 | 1 | 0.84 | 0.01 | 330 | 1.64 | 2 | 0.44 | 0.07 |
| **Researcher gender attendance** | | | | | | | | | | |
| **PwD gender** | 325 | 0.52 | 2 | 0.77 | 0.04 | 325 | 1.98 | 3 | 0.57 | 0.08 |
| **Carer gender** | 325 | 0.93 | 2 | 0.63 | 0.05 | 325 | 0.18 | 3 | 0.98 | 0.02 |
| **PwD Marital Status** | 324 | 0.72 | 2 | 0.70 | 0.05 | 324 | 1.82 | 3 | 0.61 | 0.08 |
| **Carer Marital Status** | 322 | 0.27 | 2 | 0.87 | 0.03 | 322 | 7.56 | 3 | 0.06 | 0.15 |
| **Centre** | 326 | **158.93 | 12 | <0.01 | 0.50 | 326 | **235.68 | 18 | <0.01 | 0.49 |
| **Wave** | 326 | **16.78 | 8 | 0.03 | 0.16 | 326 | **83.38 | 12 | <0.01 | 0.29 |
| **Allocation** | 326 | 1.93 | 2 | 0.38 | 0.08 | 326 | 2.73 | 3 | 0.44 | 0.09 |

**Significant at the 1% level.

Appendix 6: List of Assumption Checks for ANCOVA model

| Assumption | Check |
|---|---|
| 1. You have one dependent variable that is measured at the continuous level. | Study design - If assumption doesn't hold a different analysis model is required. |
| 2. You have one independent variable that consists of two or more categorical, independent groups. | Study design - If assumption doesn't hold a different analysis model is required. |
| 3. You have one covariate variable that is measured at the continuous level | Study design - If assumption doesn't hold a different analysis model is required. |
| 4. You should have independence of observations | Study design - If assumption doesn't hold a different analysis model is required. |
| 5. The covariate should be linearly related to the dependent variable at each level of the independent variable | Grouped scatterplot of the dependent variable against the covariate for each level of the independent variable. |
| 6. You should have homogeneity of regression slopes. | Run model including interaction term between covariate and independent variable. If significant then the assumption has been violated.<br><br>If data violates the assumption then include interaction term in the model and interpret results considering this. |
| 7. Your dependent variable should be approximately normally distributed for each group of the independent variable | Obtain Q-Q plots of the model residuals.<br><br>ANCOVA models are robust enough to deal with approximately non-normal data however if plots indicate data is very non-normal then assumption is violated and transformations to the data should be applied. |
| 8. There should be homoscedasticity | Plot of model residuals against the model predicted values for each level of the researcher attendance variable.<br><br>If assumption does not hold transformations to the data should be applied. |
| 9. There should be homogeneity of variances | Calculate the variance ratio of the independent variable. A variance ratio below 2 is sufficient to assume that the assumption holds. |
| 10. There should be no significant outliers in the groups of your independent variable in terms of the dependent variable | Evaluate the range of the residual scores. Anything above or below 3 are considered outliers.<br><br>If assumption is violated outliers can be included or excluded, however outliers within range should not be excluded. Transformations may be applied if assumption is extremely violated. |

*Source: (Grace-Martin, 2013)*

Appendix 7: Assumption of homogeneity of regression slopes results

| Model/Outcome measure | Covariate | Assumption met? | |
|---|---|---|---|
| | | Researcher Attendance analysis | Gender of Researcher in Attendance analysis |
| Participant QCPR fu1 | Participant QCPR baseline scores | No – include in model | No – include in model |
| | PwD Age | Yes | Yes |
| Participant QoL-AD fu1 | Participant QoL-AD baseline scores | No – include in model | No – include in model |
| | PwD Age | Yes | Yes |
| Participant QCPR fu2 | Participant QCPR baseline scores | No – include in model | No – include in model |
| | PwD Age | Yes | Yes |
| Participant QoL-AD fu2 | Participant QoL-AD baseline scores | Yes | Yes |
| | PwD Age | Yes | Yes |
| Carer QCPR fu1 | Carer QCPR baseline scores | Yes | Yes |
| | Carer Age | Yes | Yes |
| | PwD Age | Yes | Yes |
| Carer proxy QoL-AD fu1 | Carer proxy QoL-AD baseline scores | Yes | Yes |
| | Carer Age | Yes | Yes |
| | PwD Age | Yes | Yes |
| Carer QCPR fu2 | Carer QCPR baseline scores | Yes | Yes |
| | Carer Age | Yes | Yes |
| | PwD Age | Yes | Yes |
| Carer proxy QoL-AD fu2 | Carer proxy QoL-AD baseline scores | Yes | Yes |
| | Carer Age | Yes | Yes |
| | PwD Age | No – include in model | No – include in model |

Appendix 8: Post hoc pairwise comparison tests on Researcher attendance variable for carer QoL-AD proxy measure at follow-up 2.

| RESEARCHER ATTENDANCE | ADJUSTED VALUES | | RESEARCHER ATTENDANCE | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | *SAME RESEARCHER* | | *TWO DIFFERENT* | | *THREE DIFFERENT* | |
| | N | MEAN (SE) | MEAN DIFF SIG | EFFECT SIZE (95% CI) | MEAN DIFF SIG | EFFECT SIZE (95% CI) | MEAN DIFF SIG | EFFECT SIZE (95% CI) |
| ***QOL-AD PROXY AT FOLLOW-UP 2*** | | | | | | | | |
| ***SAME RESEARCHER*** | 114 | 31.7 (0.90) | | | 0.79 p = 0.32 | -0.09 (-0.35, 0.16) | -0.09 p = 0.93 | 0.01 (-0.28, 0.30) |
| ***TWO DIFFERENT*** | 127 | 30.9 (0.69) | -0.79 p = 0.32 | 0.09 (-0.16, 0.35) | | | -0.88 p = 0.35 | 0.11 (-0.17, 0.39) |
| ***THREE DIFFERENT*** | 79 | 31.8 (1.01) | 0.09 p = 0.93 | -0.01 (-0.30, 0.28) | 0.88 p = 0.35 | -0.11 (-0.39, 0.17) | | |

Appendix 9: Post hoc pairwise comparison test results conduced on the sensitivity analysis model significant findings for the researcher attendance variable

| OUTCOME MEASURE | RESEARCHER ATTENDANCE AT FU2 | ADJUSTED VALUES | | RESEARCHER ATTENDANCE AT FU2 | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | *Same researcher* | | *Two different* | | *Three different* | |
| | | N | Mean (SE) | MEAN DIFF SIG | EFFECT SIZE (95% CI) | MEAN DIFF SIG | EFFECT SIZE (95% CI) | MEAN DIFF SIG | EFFECT SIZE (95% CI) |
| **FOLLOW-UP 2** | | | | | | | | | |
| **PARTICIPANT QCPR** | ***Same Researcher*** | 92 | 57.9 (1.11) | | | 0.51 p = 0.64 | -0.05 (-0.34, 0.24) | -0.41 p = 0.76 | 0.04 (-0.26, 0.33) |
| | ***Two different*** | 94 | 57.4 (0.90) | -0.51 p = 0.64 | 0.05 (-0.24, 0.34) | | | -0.92 p = 0.48 | 0.09 (-0.21, 0.38) |
| | ***Three different*** | 84 | 58.3 (1.24) | 0.41 p = 0.76 | -0.04 (-0.33, 0.26) | 0.92 p = 0.48 | -0.09 (-0.38, 0.21) | | |
| **PARTICIPANT QOL-AD** | ***Same Researcher*** | 91 | 36.8 (0.87) | | | -1.30 p = 0.13 | 0.17 (-0.12, 0.46) | -0.64 p = 0.51 | 0.08 (-0.21, 0.38) |
| | ***Two different*** | 99 | 38.1 (0.70) | 1.30 p = 0.13 | -0.17 (-0.46, 0.12) | | | 0.65 p = 0.47 | -0.08 (-0.36, 0.21) |
| | ***Three different*** | 90 | 37.5 (0.88) | 0.64 p = 0.51 | -0.08 (-0.38, 0.21) | -0.65 p = 0.47 | 0.08 (-0.21, 0.36) | | |
| **CARER QOL-AD PROXY** | ***Same Researcher*** | 114 | 31.8 (0.90) | | | 1.12 p = 0.17 | -0.13 (-0.39, 0.13) | -0.24 p = 0.80 | 0.02 (-0.25, 0.29) |
| | ***Two different*** | 111 | 30.7 (0.70) | -1.12 p = 0.17 | 0.13 (-0.13, 0.39) | | | -1.35 p = 0.12 | 0.16 (-0.12, 0.43) |
| | ***Three different*** | 95 | 32.0 (0.92) | 0.24 p = 0.80 | -0.02 (-0.29, 0.25) | 1.35 p = 0.12 | -0.16 (-0.43, 0.12) | | |
| **CARER QCPR** | ***Same Researcher*** | 113 | 53.1 (1.47) | *n/a | | | | | |
| | ***Two different*** | 107 | 52.4 (1.15) | | | | | | |
| | ***Three different*** | 88 | 54.5 (1.51) | | | | | | |

*Pairwise comparisons not carried out as effect was not significant in the model

Appendix 10: Details of researcher genders and number of visits conducted by each

**Table 1:** Genders of the 43 researchers collecting data for REMCARE - individual and across study

| Researcher Number and Gender | Overall (N = 990) N (%) | At each follow-up (N = 330) | | |
|---|---|---|---|---|
| | | Baseline N (%) | Follow-up 1 N (%) | Follow-up 2 N (%) |
| ***Researcher 1 – Female*** | **1 (0.1%)** | 0 (0) | 1 (<1%) | 0 (0) |
| ***Researcher 2 – Female*** | **3 (0.3%)** | 3 (1%) | 0 (0) | 0 (0) |
| ***Researcher 3 – Female*** | **10 (1.0%)** | 7 (2%) | 3 (1%) | 0 (0) |
| ***Researcher 4 – Female*** | **27 (2.7%)** | 9 (3%) | 13 (4%) | 5 (2%) |
| ***Researcher 5 – Female*** | **1 (0.1%)** | 0 (0) | 0 (0) | 1 (<1%) |
| ***Researcher 6 – Unknown*** | **2 (0.2%)** | 1 (<1%) | 1 (<1%) | 0 (0) |
| ***Researcher 7 – Female*** | **7 (0.7%)** | 1 (<1%) | 1 (<1%) | 5 (2%) |
| ***Researcher 8 – Female*** | **51 (5.2%)** | 5 (2%) | 16 (5%) | 30 (9%) |
| ***Researcher 9 – Female*** | **2 (0.2%)** | 0 (0) | 2 (1%) | 0 (0) |
| ***Researcher 10 – Male*** | **37 (3.7%)** | 0 (0) | 11 (3%) | 26 (8%) |
| ***Researcher 11 – Unknown*** | **1 (0.1%)** | 1 (<1%) | 0 (0) | 0 (0) |
| ***Researcher 12 – Female*** | **2 (0.2%)** | 0 (0) | 2 (1%) | 0 (0) |
| ***Researcher 13 – Female*** | **63 (6.4%)** | 6 (2%) | 19 (6%) | 38 (12%) |
| ***Researcher 14 – Female*** | **45 (4.5%)** | 16 (5%) | 14 (4%) | 15 (5%) |
| ***Researcher 15 – Female*** | **14 (1.4%)** | 12 (4%) | 1 (<1%) | 1 (<1%) |
| ***Researcher 16 – Female*** | **4 (0.4%)** | 2 (<1%) | 0 (0) | 2 (1%) |
| ***Researcher 17 – Female*** | **21 (2.1%)** | 14 (4%) | 7 (2%) | 0 (0) |
| ***Researcher 18 – Female*** | **7 (0.7%)** | 5 (2%) | 1 (<1%) | 1 (<1%) |
| ***Researcher 19 – Female*** | **6 (0.6%)** | 3 (1%) | 3 (1%) | 0 (0) |
| ***Researcher 20 – Female*** | **9 (0.9%)** | 3 (1%) | 3 (1%) | 3 (1%) |
| ***Researcher 21 – Female*** | **3 (0.3%)** | 3 (1%) | 0 (0) | 0 (0) |
| ***Researcher 22 – Female*** | **18 (1.8%)** | 0 (0) | 10 (3%) | 8 (2%) |
| ***Researcher 23 – Female*** | **7 (0.7%)** | 4 (1%) | 3 (1%) | 0 (0) |
| ***Researcher 24 – Female*** | **6 (0.6%)** | 6 (2%) | 0 (0) | 0 (0) |
| ***Researcher 25 – Female*** | **8 (0.8%)** | 8 (2%) | 0 (0) | 0 (0) |
| ***Researcher 26 – Male*** | **79 (8.0%)** | 20 (6%) | 22 (7%) | 37 (11%) |
| ***Researcher 27 – Female*** | **48 (4.8%)** | 15 (5%) | 27 (8%) | 6 (2%) |
| ***Researcher 28 – Female*** | **4 (0.4%)** | 2 (1%) | 2 (1%) | 0 (0) |
| ***Researcher 29 – Male*** | **17 (1.7%)** | 5 (2%) | 12 (4%) | 0 (0) |
| ***Researcher30 – Female*** | **17 (1.7%)** | 8 (2%) | 3 (1%) | 6 (2%) |
| ***Researcher 31 – Male*** | **44 (4.4%)** | 17 (5%) | 16 (5%) | 11 (3%) |
| ***Researcher 32 – Female*** | **2 (0.2%)** | 0 (0) | 2 (1%) | 0 (0) |
| ***Researcher 33 – Male*** | **13 (1.3%)** | 4 (1%) | 2 (1%) | 7 (2%) |
| ***Researcher 34 – Female*** | **13 (1.3%)** | 7 (2%) | 5 (2%) | 1 (<1%) |
| ***Researcher 35 – Male*** | **9 (0.9%)** | 3 (1%) | 3 (1%) | 3 (1%) |
| ***Researcher 36 – Female*** | **124 (12.5%)** | 37 (12%) | 41 (12%) | 46 (14%) |
| ***Researcher 37 – Female*** | **6 (0.6%)** | 2 (1%) | 2 (1%) | 2 (1%) |
| ***Researcher 38 – Female*** | **113 (11.4%)** | 48 (15%) | 40 (12%) | 25 (8%) |
| ***Researcher 39 – Female*** | **8 (0.8%)** | 4 (1%) | 2 (1%) | 2 (1%) |
| ***Researcher 40 – Unknown*** | **1 (0.1%)** | 1 (<1%) | 0 (0) | 0 (0) |
| ***Researcher 41 – Female*** | **4 (0.4%)** | 1 (<1%) | 0 (0) | 3 (1%) |
| ***Researcher 42 – Female*** | **73 (7.4%)** | 19 (6%) | 23 (7%) | 31 (9%) |
| ***Researcher 43 – Female*** | **60 (6.1%)** | 28 (9%) | 17 (5%) | 15 (5%) |
| ***TOTAL*** | **990** | **330** | **330** | **330** |

**Table 2:** Figures of Gender of researchers across study and at each visit

| | **Genders of researchers across the study N (%)** | | | | **Number of visits conducted by each gender N (%)** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Overall** | **Baseline** | **Follow-up 1** | **Follow-up 2** | **Overall** | **Baseline** | **Follow-up 1** | **Follow-up 2** |
| *Male* | **6 (14%)** | 5 (14%) | 6 (18%) | 5 (18%) | **199 (20%)** | 45 (14%) | 64 (19%) | 77 (23%) |
| *Female* | **34 (79%)** | 28 (78%) | 27 (79%) | 23 (82%) | **787 (79%)** | 282 (86%) | 265 (80%) | 253 (77%) |
| *Not Known* | **3 (7%)** | 3 (8%) | 1 (3%) | 0 (0%) | **4 (<1%)** | 3 (<1%) | 1 (<1%) | 0 (0) |
| **Total** | **43 (100%)** | **36 (100%)** | **34 (100%)** | **28 (100%)** | **990 (100%)** | 330 (100%) | 330 (100%) | 330 (100%) |

**Researcher Gender Occurrence (N = 330)**

| | **At follow-up 1** | **At follow-up 2** |
|---|---|---|
| **Same gender researchers** | 279 (85%) | 257 (78%) |
| **Different gender researchers** | 47 (14%) | 69 (21%) |
| **Unknown** | 4 (1%) | 4 (1%) |

Appendix 11: Sensitivity Analysis results

**Table 1:** ANCOVA sensitivity results for QCPR and QoL-AD PwD data at follow-up 1 and follow-up 2

| Factor | Unknown assumed gender | DF | F-value | Significance (p-value) | DF | F-value | Significance (p-value) |
|---|---|---|---|---|---|---|---|
| | | QCPR follow-up 1 | | | QoL-AD follow-up 1 | | |
| Outcome Measure Baseline value | Female | 1 | 75.20 | **<0.01 | 1 | 97.56 | **<0.01 |
| | Male | 1 | 74.27 | **<0.01 | 1 | 96.94 | **<0.01 |
| Age | Female | 1 | 0.07 | 0.78 | 1 | 0.96 | 0.33 |
| | Male | 1 | 0.07 | 0.79 | 1 | 0.93 | 0.34 |
| Gender | Female | 1 | 0.93 | 0.34 | 1 | 2.54 | 0.11 |
| | Male | 1 | 0.98 | 0.32 | 1 | 2.49 | 0.12 |
| Marital status | Female | 1 | 0.83 | 0.36 | 1 | 0.11 | 0.74 |
| | Male | 1 | 0.78 | 0.38 | 1 | 0.11 | 0.74 |
| Centre | Female | 6 | 0.47 | 0.83 | 6 | 1.78 | 0.10 |
| | Male | 6 | 0.42 | 0.87 | 6 | 1.80 | 0.10 |
| Wave | Female | 4 | 2.46 | *0.05 | 4 | 0.56 | 0.69 |
| | Male | 4 | 2.49 | *0.04 | 4 | 0.54 | 0.70 |
| Allocation | Female | 1 | 0.18 | 0.67 | 1 | 0.68 | 0.41 |
| | Male | 1 | 0.14 | 0.71 | 1 | 0.68 | 0.41 |
| Centre x Allocation | Female | 6 | 0.16 | 0.99 | 6 | 2.09 | 0.06 |
| | Male | 6 | 0.18 | 0.98 | 6 | 2.09 | 0.06 |
| Gender of Researcher in Attendance (fu1) | Female | 2 | 3.20 | *0.05 | 2 | 5.48 | **<0.01 |
| | Male | 2 | 3.16 | *0.04 | 2 | 5.52 | **<0.01 |
| Centre x Gender of Researcher in Attendance (fu1) | Female | 8 | 0.49 | 0.86 | 9 | 0.77 | 0.65 |
| | Male | 8 | 0.78 | 0.87 | 9 | 0.76 | 0.65 |
| Baseline x Gender of Researcher in Attendance (fu1) | Female | 2 | 3.11 | *0.05 | 2 | 4.83 | **0.01 |
| | Male | 2 | 3.11 | *0.05 | 2 | 4.86 | **0.01 |
| Error (SS Within) | Female | 242 | | | 262 | | |
| | Male | 242 | | | 262 | | |
| | | QCPR follow-up 2 | | | QoL-AD follow-up 2 | | |
| Measure Baseline value | Female | 1 | 48.34 | **<0.01 | 1 | 93.59 | **<0.01 |
| | Male | 1 | 47.86 | **<0.01 | 1 | 93.02 | **<0.01 |
| Age | Female | 1 | 3.94 | *0.05 | 1 | 0.26 | 0.61 |
| | Male | 1 | 3.82 | *0.05 | 1 | 0.27 | 0.60 |
| Gender | Female | 1 | 0.005 | 0.94 | 1 | 0.19 | 0.66 |
| | Male | 1 | 0.005 | 0.94 | 1 | 0.15 | 0.70 |
| Marital status | Female | 1 | 0.19 | 0.50 | 1 | 0.70 | 0.40 |
| | Male | 1 | 0.15 | 0.70 | 1 | 0.53 | 0.47 |
| Centre | Female | 6 | 0.48 | 0.83 | 6 | 2.38 | *0.03 |
| | Male | 6 | 0.44 | 0.85 | 6 | 2.08 | 0.06 |
| Wave | Female | 4 | 0.41 | 0.80 | 4 | 0.64 | 0.63 |
| | Male | 4 | 0.42 | 0.79 | 4 | 0.70 | 0.59 |
| Allocation | Female | 1 | 0.30 | 0.59 | 1 | 0.13 | 0.72 |
| | Male | 1 | 0.28 | 0.60 | 1 | 0.15 | 0.70 |
| Centre x Allocation | Female | 6 | 0.80 | 0.57 | 6 | 0.68 | 0.67 |
| | Male | 6 | 0.81 | 0.57 | 6 | 0.67 | 0.67 |
| Gender of Researcher in Attendance (fu2) | Female | 2 | 3.97 | *0.02 | 2 | 1.17 | 0.31 |
| | Male | 2 | 3.91 | *0.02 | 2 | 0.64 | 0.53 |
| Centre x Gender of Researcher in Attendance (fu2) | Female | 9 | 1.69 | 0.09 | 9 | 0.87 | 0.53 |
| | Male | 10 | 1.52 | 0.13 | 10 | 0.85 | 0.59 |
| | Female | 2 | 3.71 | *0.03 | - | - | - |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Gender of Researcher in Attendance (fu2) x baseline** | **Male** | 2 | 3.70 | *0.03 | - | - | - |
| **Error (SS Within)** | **Female** | 235 | | | 247 | | |
| | **Male** | 234 | | | 246 | | |

**Significant at the 0.05 level. **Significant at the 0.01 level.*

**Table 2:** Adjusted means for gender of researcher in attendance groups from PwD ANCOVA sensitivity model

| OUTOME MEASURE | ASSUMED RESEARCHER GENDER | ADJUSTED VALUES AT FOLLOW-UP 2 | | | | | |
|---|---|---|---|---|---|---|---|
| | | Same Researcher Same gender | | Different Researchers Same gender | | Different Researchers Different gender | |
| | | N | Mean (SE) | N | Mean (SE) | N | Mean (SE) |
| **QCPR fu1** | **Female** | 132 | 58.0 (0.96) | 101 | 58.9 (1.15) | 43 | 55.4 (1.02) |
| | **Male** | 132 | 58.0 (0.96) | 101 | 58.8 (1.16) | 43 | 55.5 (1.03) |
| **QOL-AD FU1** | **Female** | 136 | 37.1(0.65) | 115 | 37.8 (0.82) | 46 | 38.3 (1.26) |
| | **Male** | 136 | 37.1 (0.65) | 115 | 37.7 (0.82) | 46 | 38.4 (1.26) |
| **OUTOME MEASURE** | **ASSUMED RESEARCHER GENDER** | **Same Researcher Same gender** | | **Different Researchers Same gender** | | **Different Researchers Different gender** | |
| | | **N** | **Mean (SE)** | **N** | **Mean (SE)** | **N** | **Mean (SE)** |
| **QCPR FU2** | **Female** | 92 | 58.0 (1.10) | 115 | 57.7 (1.03) | 63 | 55.2 (1.04) |
| | **Male** | 92 | 58.1 (1.10) | 115 | 57.6 (1.04) | 63 | 55.1 (1.47) |
| **QOL-AD FU2** | **Female** | 91 | 36.6 (0.86) | 122 | 37.6 (0.77) | 67 | 37.6 (0.80) |
| | **Male** | 91 | 36.6 (0.86) | 122 | 37.6 (0.77) | 67 | 37.4 (1.14) |

**Table 3:** ANCOVA sensitivity model results for carer QCPR and carer proxy QoL-AD at follow-up 1 and follow-up 2

| Factor | Unknown assumed gender | DF | F-value | Significance (p-value) | DF | F-value | Significance (p-value) |
|---|---|---|---|---|---|---|---|
| | | QCPR follow-up 1 | | | QoL-AD follow-up 1 | | |
| **Outcome Measure Baseline value** | ***Female*** | 1 | 286.4 | **<0.01 | 1 | 325.42 | **<0.01 |
| | ***Male*** | 1 | 287.19 | **<0.01 | 1 | 322.89 | **<0.01 |
| **PwD Age** | ***Female*** | 1 | 0.11 | 0.74 | 1 | 0.43 | 0.51 |
| | ***Male*** | 1 | 0.09 | 0.77 | 1 | 0.52 | 0.47 |
| **Carer Gender** | ***Female*** | 1 | 1.09 | 0.30 | 1 | 0.10 | 0.76 |
| | ***Male*** | 1 | 1.12 | 0.29 | 1 | 0.11 | 0.74 |
| **Carer Age** | ***Female*** | 1 | 0.07 | 0.79 | 1 | 0.01 | 0.98 |
| | ***Male*** | 1 | 0.06 | 0.81 | 1 | 0.01 | 0.95 |
| **PwD Gender** | ***Female*** | 1 | 4.72 | *0.03 | 1 | 0.34 | 0.56 |
| | ***Male*** | 1 | 4.86 | *0.03 | 1 | 0.24 | 0.62 |
| **Marital status** | ***Female*** | 1 | 4.62 | *0.03 | 1 | 0.81 | 0.37 |
| | ***Male*** | 1 | 4.65 | *0.03 | 1 | 0.83 | 0.36 |
| **Centre** | ***Female*** | 6 | 0.80 | 0.57 | 6 | 0.93 | 0.48 |
| | ***Male*** | 6 | 0.80 | 0.57 | 6 | 0.93 | 0.48 |
| **Wave** | ***Female*** | 4 | 0.72 | 0.58 | 4 | 1.09 | 0.36 |
| | ***Male*** | 4 | 0.74 | 0.57 | 4 | 1.08 | 0.37 |
| **Allocation** | ***Female*** | 1 | 3.89 | *0.05 | 1 | 2.24 | 0.14 |
| | ***Male*** | 1 | 3.96 | *0.05 | 1 | 2.07 | 0.15 |
| **Centre x Allocation** | ***Female*** | 6 | 1.48 | 0.19 | 6 | 2.05 | 0.06 |
| | ***Male*** | 6 | 1.47 | 0.19 | 6 | 2.22 | *0.04 |
| **Gender of Researcher in Attendance (fu1)** | ***Female*** | 2 | 0.44 | 0.65 | 2 | 1.97 | 0.14 |
| | ***Male*** | 2 | 0.41 | 0.67 | 2 | 2.06 | 0.13 |
| **Centre x Gender of Researcher in Attendance (fu1)** | ***Female*** | 9 | 1.35 | 0.21 | 9 | 0.79 | 0.63 |
| | ***Male*** | 9 | 1.40 | 0.19 | 9 | 0.82 | 0.60 |
| **Error (SS Within)** | ***Female*** | 266 | | | 281 | | |
| | ***Male*** | 266 | | | 281 | | |
| | | QCPR follow-up 2 | | | QoL-AD follow-up 2 | | |
| **Outcome Measure Baseline value** | ***Female*** | 1 | 191.09 | **<0.01 | 1 | 207.88 | **<0.01 |
| | ***Male*** | 1 | 190.44 | **<0.01 | 1 | 210.19 | **<0.01 |
| **PwD Age** | ***Female*** | 1 | 0.91 | 0.34 | 1 | 0.08 | 0.77 |
| | ***Male*** | 1 | 1.24 | 0.27 | 1 | 0.32 | 0.57 |
| **Carer Gender** | ***Female*** | 1 | 0.61 | 0.44 | 1 | 2.71 | 0.10 |
| | ***Male*** | 1 | 0.77 | 0.38 | 1 | 3.55 | 0.06 |
| **Carer Age** | ***Female*** | 1 | 1.74 | 0.19 | 1 | 0.83 | 0.36 |
| | ***Male*** | 1 | 2.28 | 0.13 | 1 | 1.67 | 0.20 |
| **PwD Gender** | ***Female*** | 1 | 0.01 | 0.93 | 1 | 0.56 | 0.46 |
| | ***Male*** | 1 | 0.05 | 0.82 | 1 | 0.89 | 0.35 |
| **Marital status** | ***Female*** | 1 | 0.01 | 0.96 | 1 | 0.77 | 0.38 |
| | ***Male*** | 1 | 0.02 | 0.89 | 1 | 0.30 | 0.58 |
| **Centre** | ***Female*** | 6 | 1.23 | 0.28 | 6 | 2.06 | 0.06 |
| | ***Male*** | 6 | 1.23 | 0.29 | 6 | 3.58 | **0.01 |
| **Wave** | ***Female*** | 4 | 0.36 | 0.84 | 4 | 0.97 | 0.43 |
| | ***Male*** | 4 | 0.47 | 0.76 | 4 | 1.02 | 0.40 |
| **Allocation** | ***Female*** | 1 | 0.06 | 0.81 | 1 | 0.58 | 0.45 |
| | ***Male*** | 1 | 0.05 | 0.83 | 1 | 0.70 | 0.40 |
| **Centre x Allocation** | ***Female*** | 6 | 1.11 | 0.36 | 6 | 0.49 | 0.81 |
| | ***Male*** | 6 | 1.10 | 0.36 | 6 | 0.58 | 0.75 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Gender of Researcher in Attendance (fu2)** | ***Female*** | 2 | 0.35 | 0.71 | 2 | 4.18 | *0.02 |
| | ***Male*** | 2 | 0.38 | 0.68 | 2 | 3.98 | *0.02 |
| **Centre x Gender of Researcher in Attendance (fu2)** | ***Female*** | 10 | 1.27 | 0.25 | 10 | 1.24 | 0.26 |
| | ***Male*** | 10 | 1.26 | 0.25 | 10 | 2.10 | *0.03 |
| **PwD Age x Gender of Researcher in Attendance** | ***Female*** | - | - | - | 2 | 3.79 | *0.02 |
| | ***Male*** | - | - | - | 2 | 3.93 | *0.02 |
| **Error (SS Within)** | ***Female*** | 272 | | | 282 | | |
| | ***Male*** | 272 | | | 282 | | |

**Table 4:** Adjusted means for gender of researcher in attendance groups from carer ANCOVA sensitivity model

| OUTOME MEASURE | ASSUMED RESEARCHER GENDER | ADJUSTED VALUES AT FOLLOW-UP 2 | | | | | |
|---|---|---|---|---|---|---|---|
| | | Same Researcher Same gender | | Different Researchers Same gender | | Different Researchers Different gender | |
| | | N | Mean (SE) | N | Mean (SE) | N | Mean (SE) |
| **QCPR fu1** | **Female** | 145 | 52.6 (0.99) | 108 | 51.8 (1.11) | 48 | 52.5 (1.78) |
| | **Male** | 145 | 52.6 (0.99) | 108 | 51.8 (1.11) | 48 | 52.5 (1.78) |
| **QOL-AD FU1** | **Female** | 153 | 31.9 (0.67) | 115 | 30.6 (0.76) | 48 | 31.6 (1.23) |
| | **Male** | 153 | 31.9 (0.67) | 115 | 30.5 (0.75) | 48 | 31.6 (1.22) |
| **OUTOME MEASURE** | **ASSUMED RESEARCHER GENDER** | **Same Researcher Same gender** | | **Different Researchers Same gender** | | **Different Researchers Different gender** | |
| | | **N** | **Mean (SE)** | **N** | **Mean (SE)** | **N** | **Mean (SE)** |
| **QCPR FU2** | **Female** | 113 | 53.3 (1.46) | 127 | 52.2 (1.28) | 68 | 52.0 (1.92) |
| | **Male** | 113 | 53.2 (1.46) | 126 | 52.1 (1.28) | 69 | 52.5 (1.58) |
| **QOL-AD FU2** | **Female** | 114 | 31.9 (0.90) | 136 | 30.6 (0.78) | 70 | 31.3 (1.18) |
| | **Male** | 114 | 31.8 (0.88) | 135 | 30.5 (0.77) | 71 | 32.9 (0.95) |