Insensitivity to the effectiveness of talking-therapy: the impact of the baseline-risk

Florian Vogt

June 2014

Submitted in partial fulfilment of the requirements for the degree of Doctor in Clinical Psychology (DClinPsy), Royal Holloway, University of London

Acknowledgements

Foremost, I thank my wife Ariadna and my children Magali and Nina Maya for supporting me in carrying out this project and showing admirable tolerance with me for the times in which I failed to take part, contribute, and play. I also thank Dr Abigail Wroe for supervising me on this project and providing expert advice on the development and execution of the study. Finally, I would like to thank the participants who volunteered their time and ultimately made this study possible.

Table of contents

Table of Figures

Abstract

Objectives. People suffering from mental health conditions, often do not seek professional help. One of the reasons for this is that they do not consider talking-therapies sufficiently effective. It has been shown that among physical health conditions the rate at which people recover by themselves from a condition, as compared to those who do not (i.e. baseline-risk), unduly influences how effective people judge respective treatments. Treatments for conditions from which many people recover by themselves are considered as more effective than they actually are, as people credit the treatments for those that have recovered by themselves; the reverse is true for conditions from which many people do not recover by themselves. People may judge talking-therapy on the basis of the baseline-risk, to the detriment of the actual treatment effect of talking-therapy, conceptualised as the absolute or relative reduction of risk.

Design. A mixed factorial within-between subjects experiment.

Participants. A general population sample (N=202), in which 75.8% of participants had symptoms of depression and anxiety indicating a mental health condition.

Methods. Participants took part in a web-based experiment during which they were shown six vignettes about common mental health conditions and the effect of talking-therapy on these compared to no treatment. The six vignettes varied in the baseline-risk (high vs. low), the absolute risk reduction (high vs. low), and the relative risk reduction (high vs. low). The dependent variable was

the perceived effectiveness of talking-therapy, measured on a visual analogue scale from 0 (ineffective) to 100 (extremely effective). Analysis of covariance was conducted to control for potential confounding variables, including numeracy and mood.

Results. Talking-therapies were judged on the basis of the baseline-risk of the condition. A higher baseline-risk was associated with a lower perceived effectiveness. This impact of the baseline-risk was not moderated by numeracy or mood. Talking-therapies were also judged on the basis of the absolute risk reduction but not by the relative risk reduction. There was some evidence that those with lower numeracy were less sensitive to differences in the treatment effect, that is, their judgments of effectiveness decreased less with a decreasing treatment effect.

Conclusion. The evidence suggests that the effectiveness of talking-therapy is influenced by the baseline-risk of common mental health conditions. To address people's judgement that talking-therapy is insufficiently effective may require considering the negative impact of the high baseline-risk inherent to common mental health conditions. In turn, this may increase uptake of talking-therapy for common mental health conditions.

1. Introduction

Only a quarter of people with a mental health condition are in treatment (Mental Health Policy Group, 2012). One reason for this is that many people suffering from mental health conditions, even if severe, do not seek professional help (Oliver et al, 2005). There are several reasons for the low rates of help-seeking for mental health conditions among adults in the general population, including the perception that treatment will not help (Meltzer, 2000; Stecker, Fortney, Hamilton, Sherbourne, & Ajzen, 2010).

Research into decision making about treatments for physical health conditions has demonstrated that a treatment is less likely to be perceived as helping when the natural recovery rate for the condition in the absence of treatment is low (Vogt, Mason, & Marteau, 2012). In other words, a treatment for a physical health condition from which few people would recover by themselves, is unduly judged as less effective. The inverse of the natural recovery rate of a condition or the proportion of those who do not get better by themselves is often referred to as the baseline-risk (B_RISK).

Many people with common mental health conditions (CMHCs) do not get better by themselves (i.e. in the absence of treatment) (e.g. Bisson & Andrew, 2007; Hunot, Churchill, Teixeira, & Silva de Lima, 2007). For example, 89% of people with Post Traumatic Stress Disorder (PTSD; Bisson & Andrew, 2007), 86% with Generalised Anxiety Disorder (GAD; Hunot et al., 2007), 70% of people with depression (Layard, Clark, Knapp, & Mayraz, 2007), 95% of people with phobia (Layard et al., 2007), and 95% of people with obsessive-compulsive disorder

(OCD; Layard et al., 2007) do not get better by themselves. Thus, if perceptions of treatments for mental health conditions are judged on the basis of the natural recovery rate in the absence of treatment, they are likely to be judged as less effective than they actually are. In turn, this may contribute to the low rates of seeking professional help. Information aimed at increasing help-seeking could then be strengthened by taking into account the impact of those that do not get better in the absence of treatment, the B_RISK of CMHCs.

The current research aims to address the question of whether perceptions of treatments for mental health conditions are judged on the basis of the B_RISK.

1.1. The human and financial cost of mental health problems

In the UK about 16% of adults are experiencing a CMHC (McManus, 2009; Mental Health Policy Group, 2012). CMHCs are mental health conditions that cause marked emotional distress and interfere with daily function, but do not usually affect insight or cognition. The term CMHC is used throughout this thesis as is it used by the NHS in online information (NHS Choices, 2014). Other terms, such as Common Mental Health Problems or Disorders, are used interchangeably and describe the same difficulties (NICE, 2011a). CMHCs comprise different types of mood conditions. Research suggests that more than half of people with a CMHC, have a mixed anxiety and depressive disorder, other common ones are depression, obsessive compulsive disorder (OCD), generalised anxiety disorder (GAD), post-traumatic stress disorder (PTSD), panic disorders, social anxiety disorder, and phobias (McManus, 2009).

One way of assessing the impact of CMHCs is by judging the disability that the conditions place upon the individual. The World Health Organisation (WHO) measures this using disability-adjusted life years (DALYs) (WHO, 2014). This time-based measure combines years of life lost due to premature mortality and years of life lost due to time lived in states of less than full health. It allows comparison of disability across the spectrum of diseases, injuries, and risk factors, such as smoking. The WHO reports that on its own, depression places a huge burden on the individual (Mathers, Fat, & Boerma, 2008). In terms of DALYs, it ranks third place world-wide, eighth place in low-income countries, but first place in middle- and high-income countries. Only lower-respiratory infections and diarrhoeal diseases have a larger burden of disease world-wide.

Another way of looking at the impact of CMHCs is to look at suffering, as measured by quality of life. In other words, how satisfying an individual perceives his/her life. Layard and colleagues show that poor mental health, measured six years previously, contributes more to current quality of life than a person's current physical health or household income (Mental Health Policy Group, 2012).

There is also a huge economic cost for society attached to CMHCs. Depression and anxiety make it much more difficult to work, and even those who are working have high rates of sickness absence leading to a large financial burden (Layard, 2005). In the Layard report, CMHCs are calculated to lead to annual losses of output to the size of £17bn (€24bn, $30bn), or 1.5% of UK gross domestic product.

1.2. The benefit of talking-therapy

A solution to counter the impact of CMHCs is available in form of treatment. Since the 1950s, psychoactive medications are used to help with mental health conditions (Layard, 2012). Since the 1970s, psychological or talking-therapies have become more common, the most studied of which is cognitive behaviour therapy (CBT) (Layard, 2012). Both varieties of treatment have been tested in hundreds of trials and the results analysed and summarised in several Cochrane and National Institute for Health and Clinical Excellence (NICE) reviews (Hunot et al., 2007; NICE, 2009, 2011b). Psychoactive medication and talking-therapy, including CBT and some other forms, such as Interpersonal Psychotherapy, are found to be effective and provide substantial improvements or cure for many patients. As a result, the NICE Guidelines recommend that both medication and talking-therapy should be available as options for treating CMHCs (NICE, 2009, 2011b). While the evidence suggests that both medication and talking-therapy have similar recovery rates, the effect of talking-therapy appears to be more long-lasting (Layard et al., 2007; Spielmans, Berman, & Usitalo, 2011).

Research has suggested that, not only is talking-therapy effective, it is also cost-effective. In terms of Quality Adjusted Life Years (QALYs), talking-therapy compares favourably to the recommended maximum cost of treatment set by NICE at £30,000. Talking-therapy for CMHCs also compares favourably to many common treatments for physical illnesses. The cost per additional QUALY of CBT for depression is £6,700; by comparison, statins for cardio-

vascular disease can cost £14,000 and topirimate for epilepsy costs £900 (Mental Health Policy Group, 2012).

The costs for treating CMHCs are thought to be more than recovered by savings made to the Department of Work and Pensions and the Her Majesty's Revenue and Customs (HMRC) in reduced benefits and additional taxes as more people work (Mental Health Policy Group, 2012). Large savings are also likely to be made directly within the NHS because of reduced physical health-care costs. Evidence from a meta-analysis shows that physical health-care costs for patients receiving talking-therapy for CMHCs were lower compared to those not receiving such therapy (Chiles, Lambert, & Hatch, 1999). The lower costs were such that they more than covered the costs of the additional talking-therapy, in essence reducing the total health-care-costs.

1.3. The low uptake of talking-therapy

Despite the existence and availability of effective treatments for CMHCs, uptake is low. Internationally, many people with CMHCs are not being treated (Alonso et al., 2007). The situation is similar in the UK, where only a quarter (24%) of people with a CMHC are in receipt of treatment (McManus, 2009). Most individuals in the UK receive some form of medication, with 14% of adults with such conditions taking psychoactive medication. Only 5% receive talking-therapy and 5% receive both, medication and talking-therapy.

In response to this concern, and the economic drive to treat CMHCs, Improving Access to Psychological Therapies (IAPT) was launched in 2008 following a report by Layard (Mental Health Policy Group, 2006) with the ambition for psychological services to be more available. The aim of IAPT was providing

evidence based talking-therapy for anxiety and depression to 15% of those suffering from these conditions by 2014.

1.4. Attempts to increase uptake

Despite the enormity of the problem, a recent review of interventions to increase help-seeking for CMHCs identified only six peer-reviewed interventions of good quality on the topic (e.g. randomised) (Gulliver, Griffiths, Christensen, & Brewer, 2012). The interventions that were reviewed targeted factors such as (i) knowledge about mental health conditions to help recognise, manage, or prevent them, (ii) beliefs aimed at reducing stigma, or (iii) provided information about where to find and how to access providers. Whereas most interventions achieved changes in knowledge or beliefs consistent with seeking help, only one of three interventions that measured help-seeking increased behaviour. The review concluded that more work is needed to develop interventions that change help-seeking behaviour.

As described, to increase uptake of talking-therapies in the UK, the nationwide IAPT service was introduced (Layard, 2005). It has had an impact on those seeking help, but cuts are now seen to the IAPT budgets in a number of primary care trusts, warranting some concern about the future impact (Mental Health Policy Group, 2012).

1.5. What explains the low uptake

The use of mental health treatment sits within the realm of help-seeking behaviour. In the health context, help-seeking is used to describe an adaptive form of coping with illness including the seeking of help from professionals

(Mechanic, 1962). Help-seeking developed in the literature on illness behaviour, which described the way people monitor their bodies, define and interpret their symptoms, take preventive or remedial action, or utilize the health care system (Mechanic, 1962). As far back as in the 1970s, only 10% of people consulted professionals for significant health symptoms they experienced (Tuckett, 1976).

The reasons put forward for these low rates of help-seeking in the health-context in the 1970s included a change in the nature of health conditions. Chronic illness and mental health conditions became major health concerns in the 20th century, replacing acute conditions as the main service provided by the health care system. Such conditions can be difficult to identify and interpret as something that is appropriate for professional health care. It has therefore been argued that the process of help-seeking has become more about the individual's decision making, than about the condition's acute impact on functioning (Rosenstock & Kirscht, 1979).

1.5.1. Psychological models

Four levels have been described that can help understand health care utilisation: the individual, health care groups or teams, organisations providing health care, and the larger health care system in which individual organizations are embedded (Ferlie & Shortell, 2001). These levels were already discussed in the context of help-seeking behaviour in 1976: including investigations of the individual's disposition to use certain kinds of services, patients' interpretations of their symptoms, and research on the effect of the organization and the issue of accessibility (Lewis, Fein, & Mechanic, 1976). Different levels of explanation

require different types of theory. For example, psychological theories will be more relevant for understanding individuals and teams, and theories of organisational change will be more relevant to hospitals or trusts (Walker et al., 2003).

Despite the wide range of factors that may influence the use of professional health care, even the least individually based theories which consider that the use of professional health care is determined by organisational factors assume that change is channelled through the actions of individuals (Iles & Sutherland, 2001). This is one reason why psychological theories have been influential in explaining health care behaviours (Walker et al., 2003).

Another is that they are thought to be more amenable to change. It is widely accepted that it is the attribution of events rather than the events themselves that motivate behaviour (Bandura, 1986; Beck, Rush, Shaw, & Emery, 1979). For example, research on health behaviours showed how the beliefs about the causes and controllability of their illness impacted on choice of treatment (Bradley et al., 1987). The relative importance of cognitive factors in determining health behaviours has led to the development of a large number of theories of health behaviour, which collectively are often referred to as social cognition models. Consistent with the hypothesis that it is an individual's attribution of events rather than the events themselves, these models postulate that cognitions about the behaviour and the illness meditate the impact of other factors (e.g. sociodemographic, health status, personality traits, genes) on the behaviour (Conner & Norman, 1996).

Evidence from meta-analyses supports the so called social cognition models. They show that they account for a substantial proportion of the variance in behaviour, leaving little or no variance that is explained by other factors, apart from past behaviour, which often accounts for unique variance (Floyd, Prentice-Dunn, & Rogers, 2000; Godin & Kok, 1996; Milne, Sheeran, & Orbell, 2000).

Two models will be explored in more detail to illustrate how the low rate of seeking help for CMHCs may be explained: the Health Belief Model (HBM) and the Theory of Planned Behaviour (TPB). The HBM is described because it was designed specifically to investigate uptake of health care programs and remains one of the most commonly used theories in health education and health promotion (Glanz, Rimer, & Lewis, 2002; Painter, Borba, Hynes, Mays, & Glanz, 2008), including mental health care use (Henshaw & Freedman-Doan, 2009). The TPB was selected because over the last three decades it has been considered to be the dominant theoretical approach to guide research on health-related behaviour (Sniehotta, Presseau, & Araújo-Soares, 2014).

Health belief model

One of the first social cognition models that looked at individual decision making with regards to health behaviours was the HBM (Becker, Maiman, Kirscht, Haefner, & Drachman, 1977; Rosenstock, Strecher, & Becker, 1994). This model was initially proposed by health researchers involved in promoting health care programs and it attempts to understand the beliefs that make a behaviour more or less attractive. According to the HBM, the likelihood that individuals will take action, i.e. perform some health, illness or sick-role behaviour, is dependent on the outcome of two assessments they make. One of the assessments pertains to the threat of a health problem, and the other

weighs the pros and cons of taking action (also referred to as the behavioural evaluation). The factors included in the evaluation of threat are:

- ‘Perceived severity of the health problem’: People consider how severe the organic, financial, psychological, and social consequences are likely to be if they develop the problem or leave it untreated. The more serious they believe the effects will be, the more likely they are to take action. For example, the degree to which people believe that OCD has a severe impact on their ability to function normally in everyday life, should influence who likely they are to seek professional help.

- ‘Perceived susceptibility to the health problem’: Individuals evaluate the likelihood of developing/having the problem. The more vulnerable they perceive themselves to be, the more likely they will take action. For example, accepting that one has OCD, should make it more likely to seek professional help.

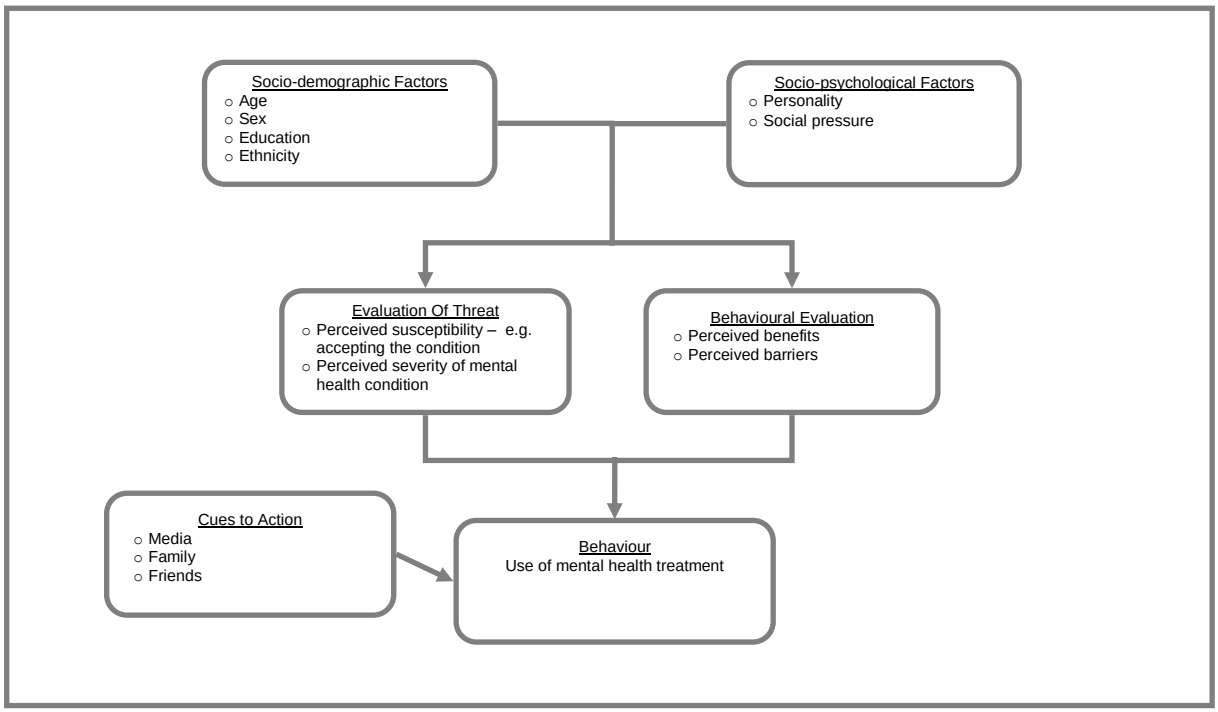
The factors that are included in the behavioural evaluation are:

- ‘Perceived benefits’: Individuals also evaluate the potential advantages of initiating a particular health behaviour or stopping some harmful health behaviour. The greater the benefits are perceived the more likely an individual will engage in the health behaviour. For instance, an individual who is deciding to start seeing a therapist might think ‘going to have therapy will help me with my OCD’.

- ‘Perceived barriers’: People often assess the costs of altering a particular behaviour. The greater the costs are perceived to be, the less likely an individual will engage in the behaviour. For instance, an individual who has

decided to go to a therapist, might realise that going to see a therapist will take a lot of time, and thus decide against it.

Figure 1.1: Conceptualisation of the Health Belief Model



Note: Adapted from Rosenstock et al., 1994

In addition to these four factors, three others are believed to modify people's decision to seek help. These factors are socio-demographic variables, socio-psychological variables (e.g. personality traits, social pressure), which are regarded to influence the evaluation of threat and the behavioural evaluation, and cues to action (i.e. people who are reminded about a potential health problem are more likely to act). Thus, for example, individuals whose close siblings have developed OCD are likely to perceive a greater threat of illness than individuals whose siblings are in good health. Similarly, it might be a trigger to seek help if the sibling's mental health declines.

The HBM has been applied to a broad range of health behaviours and populations (Harrison, Mullen, & Green, 1992; Sheeran & Abraham, 1996).

Three main areas of behaviours have been identified, including (i) preventive health behaviours (e.g. stopping smoking), (ii) compliance behaviours (e.g. following medical regimes after professional diagnoses, and (iii) clinic use, that is visiting health professionals for a variety of reasons. A meta-analysis of 234 studies looked at the predictive ability of the HBM components (Harrison et al., 1992). The results showed that each of the four components, susceptibility, severity, benefits and barriers were predictive of behaviour, with each component accounting for approximately 4% of variance in behaviour across studies. A more recent meta-analysis that assessed whether the components could longitudinally predict behaviour found that benefits and barriers were the strongest predictors, predicting 7% and 9%, respectively, in behaviour (Carpenter, 2010).

Theory of Planed Behaviour

Another model is the TPB which originated in the field of social psychology (Ajzen, 1991). It was originally developed from early work on the psychological processes demonstrating that attitudes might cause behaviour and the failure to predict behaviour from knowledge (Fishbein, 1963). The work led to a new variable between attitudes and behaviour, behavioural intention, which showed to be a powerful explanatory factor (Ajzen & Fishbein, 2005). The model includes four main components:

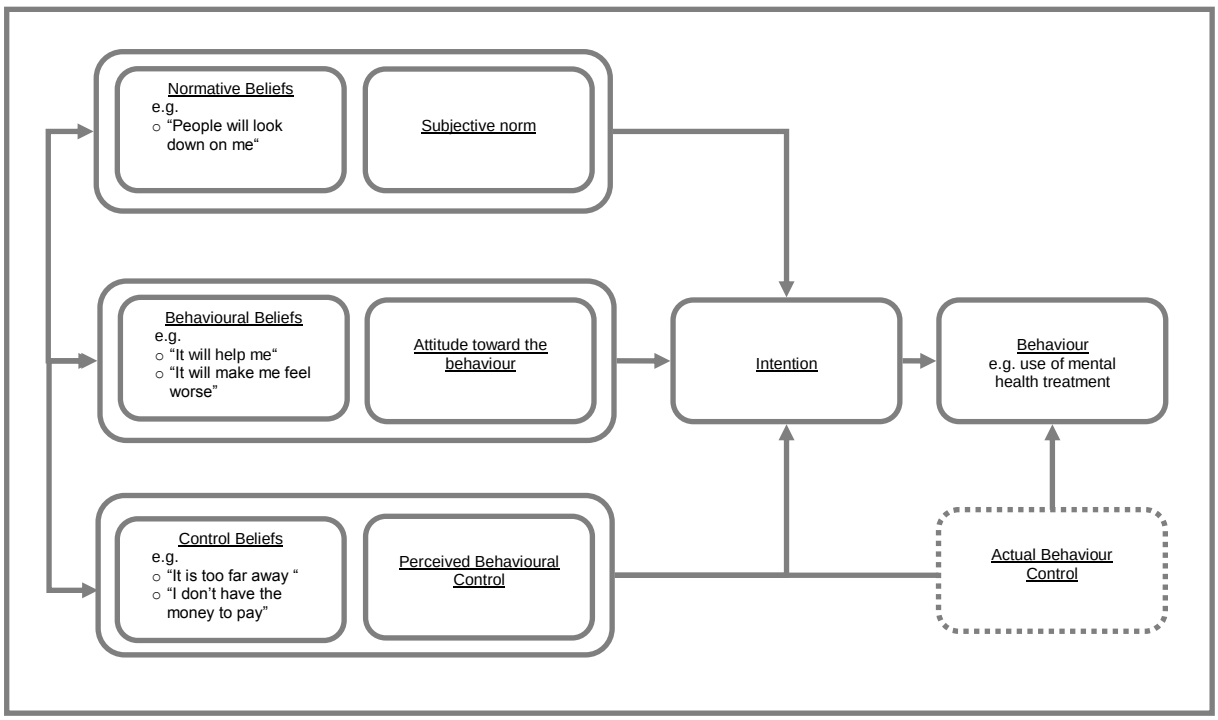
- ‘Behavioural intention’: Behavioural intention is defined as a conscious plan to perform a behaviour.

- ‘Perceived behavioural control’: Perceived behavioural control refers to people's perceptions of their ability to perform a given behaviour.

- ‘Attitude’: Attitude towards a behaviour can be defined as an overall positive or negative evaluation of engaging in a behaviour.

- ‘Subjective norm’: Subjective norm can be described as an individual’s perception of social normative pressures, or relevant others’ beliefs that he or she should or should not perform a behaviour.

Figure 1.2: Conceptualisation of the Theory of Planned Behaviour



Note: Adapted from Ajzen, 1991

The TPB posits that an individual’s intention to engage in a behaviour, and the degree of perceived behavioural control over performing the behaviour are the proximal determinants of behaviour (Ajzen, 1991). An individual's intention to engage in a behaviour is determined by three factors: the attitude towards the behaviour, subjective norm and perceived behavioural control over it. The attitude towards a behaviour is proposed to be determined by the combination of beliefs about the consequences of the behaviour (e.g. it will reduce my

symptoms, it will be expensive) and evaluations of those consequences (e.g. this is desirable vs. undesirable). The subjective norm is based on the combination of the perceptions of the views of other individuals about the behaviour (e.g. my friend would approve of seeking-help), and the strength of the individual's desire to gain approval of these individuals (e.g. it is important to me what my friend thinks). Perceived behavioural control is determined by the combination of beliefs about the perceived likelihood of potential barriers and facilitators to performing the behaviour (e.g. there is a long waiting time, the service is far away) and the perceived power of these factors to inhibit or facilitate the behaviour. It is by measuring these underlying beliefs that the TPB becomes sensitive to diverse issues such as accessibility of services, patient preferences, or stigma.

Several meta-analyses have reviewed this model (Armitage & Conner, 2001; Godin & Kok, 1996). Godin and Kok (1996) focussed on health-related behaviours and found an average correlation of r = .46 between attitude and intention, r = .34 between subjective norm and intention, and r = .46 between perceived behavioural control and intention. Attitude, social norm, and perceived behavioural control together explained 41% of variance in intention, whilst intention and perceived behavioural control explained 34% of variance in behaviour.

1.5.2. Review of predictors of uptake

Psychological models have been popular in accounting for the low uptake of treatments for CMHCs. Models that have been used to account for decisions to engage in treatment for mental health conditions, include the HBM (e.g.

Connelly, 1984; Kelly, Mamon, & Scott, 1987; Pan & Tantam, 1989), the TPB (e.g. Hyland, Boduszek, Shevlin, & Adamson, 2012; J. P. Smith, Tran, & Thompson, 2008; Stecker et al., 2010; Westerhof, Maessen, de Bruijn, & Smets, 2008), the service utilization framework (e.g. Aday & Andersen, 1974; Andrews, Issakidis, & Carter, 2001), the stages of help-seeking model (Rickwood, Deane, Wilson, & Ciarrochi, 2005), and the network episode model (Pescosolido, Gardner, & Lubell, 1998; Pescosolido, 1992).

A systematic review looking at the use of theoretical frameworks in studying help-seeking behaviour for CMHCs found that the majority of studies applied no framework (i.e. 81%) (Rickwood & Thomas, 2012) and that the most common framework was the TPB (Ajzen, 1991). Most models consider factors such as motivation, attitude, perceived threat, and beliefs about enabling resources and beliefs about the benefits of therapy as predictors of help-seeking behaviour. Indeed, a review of the different models concludes that there is substantial overlap (Fishbein et al., 2001).

Another review investigated the most frequently reported perceived barriers and facilitators to help-seeking behaviour for CMHCs (Gulliver, Griffiths, & Christensen, 2010). In the seven studies that the review included, the most frequently mentioned barriers were being uncomfortable talking about personal things, a belief that self-reliance was preferable to seeking help, the belief that support would not be beneficial, fear of stigma, and that it was too expensive. For example, in one study conducted among over 1000 adolescents, 55% of respondents believed that no person or service could help them with their CMHC and that 53% believed that their problem was too personal to tell anyone (Dubow, Lovko Jr., & Kausch, 1990).

While the frequency with which a particular belief is held in the population of people with and without a CMHC is helpful at providing an overview of the negative and positive views held, this information alone does not indicate the extent to which a respective belief is important for people's decisions to seek help. That is, although some beliefs may be held very strongly, they may not actually determine a persons' motivation to seek help or seeking help itself. To establish the relative importance of beliefs with regards to help seeking behaviour it is necessary to consider the association that beliefs have with seeking help for a CMHC, or a proxy thereof (i.e. intention to seek help) if actual behaviour was not measured. Unfortunately, no review could be identified that investigated the predictors of help-seeking behaviour in the context of CMHCs.

The literature shows that several studies could predict the intention to seek-help for CMHCs (Hyland, Boduszek, et al., 2012; Hyland, McLaughlin, Boduszek, & Prentice, 2012; Mackenzie, Gekoski, & Knox, 2006; Mackenzie, Knox, Gekoski, & Macaulay, 2004; Mills, 2010; Skogstad, Deane, & Spicer, 2006; J. P. Smith et al., 2008; Stecker et al., 2010; Vogel, Wester, Wei, & Boysen, 2005; Westerhof et al., 2008). In studies that used multivariate approaches, the variance accounted for in intention ranged between 30% to 92%. This is broadly consistent with the findings reported for the TPB in predicting behavioural intention across other health behaviours.

While some studies only used attitude to predict intention (Mackenzie et al., 2006, 2004; J. P. Smith et al., 2008; Westerhof et al., 2008), others included other factors from the TPB, by also measuring and evaluating the explanatory capacities of social norm and perceived behavioural control (Hyland,

Boduszek, et al., 2012; Hyland, McLaughlin, et al., 2012; Mills, 2010; Skogstad et al., 2006; Stecker et al., 2010; Vogel et al., 2005). When comparing the different factors, attitude emerged as the strongest predictor of intention in three studies (Mills, 2010; Skogstad et al., 2006; Vogel et al., 2005), followed by social norm in two studies (Hyland, Boduszek, et al., 2012; Skogstad et al., 2006), and perceived behavioural control in one study (Hyland, McLaughlin, et al., 2012).

Some studies have also looked more closely at the importance of individual beliefs, such as those proposed to underlie the attitude or social norm. For example, one study examined which beliefs are most predictive of attitude (Vogel et al., 2005). It found that attitude was most predicted by beliefs about treatment effectiveness, as compared to stigma, treatment fear, and fear of disclosure (Vogel et al., 2005). Other studies examined the extent to which beliefs directly predict intention (Mackenzie et al., 2006, 2004; Stecker et al., 2010; Westerhof et al., 2008). In one of these, beliefs about treatment effectiveness were the most predictive of intention to seek help (Stecker et al., 2010). The others looked at the predictive ability of clusters of beliefs, described as psychological openness ("the extent to which individuals are open to acknowledging psychological problems and to the possibility of seeking professional help for them" (Mackenzie et al., 2004: p 2420), help-seeking propensity (i.e. "the extent to which individuals believe they are willing and able to seek professional psychological help" (Mackenzie et al., 2004: p 2420), and indifference to stigma. Two found help-seeking propensity (Mackenzie et al., 2006, 2004) and one psychological openness (Westerhof et al., 2008) most predictive of the intention to seek help. Unfortunately, none of these latter three

studies measured beliefs about treatment effectiveness. A narrative review comparing the relative importance of stigma versus the belief that seeking help is beneficial concluded that beliefs about treatment effectiveness are at the core of the intention to seek help for CMHCs (Schomerus & Angermeyer, 2008).

Only two studies were identified that predicted actual behaviour (Stecker et al., 2010; Vogel et al., 2005). These used a prospective design, that is, they assessed the predictors at a time prior to measuring the dependent variable. Such a design strengthens the assumptions about the causal relations implied in many cross-sectional studies. In one of these two studies, intention explained 29% of the variance in seeking help for CMHCs and furthermore showed that beliefs about treatment effectiveness were associated with seeking help, when other beliefs were not (Stecker et al., 2010). In the other, beliefs about the effectiveness and having a distressing experience predicted seeking help, while controlling for gender, risk of treatment, and comfort of self-disclosure, all of which had no impact on help-seeking behaviour (Vogel et al., 2005).

The literature on the predictors of help-seeking for CMHCs revealed some weaknesses. For example, few studies predicted actual behaviour and most investigated intention. Many studies asserting to use specific theoretical frameworks, such as the TPB, did not follow the theory's framework of how to measure the constructs (Mackenzie et al., 2006, 2004; Skogstad et al., 2006; Vogel et al., 2005; Westerhof et al., 2008). For example, some created clusters of beliefs that appear to overlap with intention (e.g. help-seeking propensity; Mackenzie et al., 2006, 2004; Westerhof et al., 2008). Also, some studies were

not specific to talking-therapy, but described treatment for CMHCs as a combination of talking-therapy and medication, or as seeking out a mental health professional (Mackenzie et al., 2006; Skogstad et al., 2006; Stecker et al., 2010). However, others were specific to talking-therapy (e.g. Hyland, Boduszek, et al., 2012; Hyland, McLaughlin, et al., 2012; Mackenzie et al., 2004; Vogel et al., 2005; Westerhof et al., 2008).

In summary, factors postulated by social cognition models, appear to be a valid and useful framework to help understand the low rates of help-seeking for CMHCs. These factors, account for a substantial amount of variance in intention and behaviour. Bearing in mind the small number of studies, the attitude, more so than the subjective norm or the perceived behavioural control, predicted help-seeking. A key belief driving this relationship seemed to be beliefs about the effectiveness of the treatment at reducing symptoms. Public or private acceptability of seeking-help, stigma, or accessibility were other, albeit seemingly less important predictors. The findings are consistent with Layard's analysis of the UK (Layard, 2013). While Layard did not conduct any primary research and the sources for his conclusions are not referenced, he postulated that people not realising that treatment is effective, is one of three factors that account for low treatment use. The other two factors postulated by Layard are: people and their relations are ashamed to admit there is a problem, and that facilities are not available.

1.6. The neglected predictor: beliefs about the effectiveness

The analysis of the existing literature on the predictors of help-seeking behaviours thus suggests that low use of talking-therapies for CMHCs is

associated with beliefs that treatment is ineffective, normative beliefs related to stigma, and the inaccessibility of services. It would follow then that an intervention to increase uptake should target such issues.

In the UK, IAPT focussed on one of these factors, namely, the accessibility of talking-therapies. However, none of the interventions in the review of interventions aimed at increasing help-seeking for CMHCs targeted beliefs about the effectiveness of treatment (Gulliver et al., 2012). This is surprising given that the evidence suggested that this is among the strongest and most consistent individual predictors of help-seeking behaviour (Stecker et al., 2010; Vogel et al., 2005). Potentially, this is a result of the complexity of how beliefs that treatment is not effective might be addressed.

Ultimately, even in the UK, addressing low rates of help-seeking by increasing only the accessibility will be limited to helping people with CMHCs whose main barrier is being unable to access a service. Increasing accessibility alone is unlikely to entice those who do not believe that talking-therapy is effective or those who avoid treatment for fear of stigma. Policy and interventions are likely to be more efficient if they addressed all factors underlying low rates of help-seeking.

That beliefs about the effectiveness are a key determinant of treatment for CMHCs is not unusual. Beliefs about the effectiveness have demonstrated to predict treatment use in a variety of contexts, for example stop smoking treatment or asthma medication (Floyd et al., 2000; Hammond, McDonald, Fong, & Borland, 2004; Horne & Weinman, 1998; Weinstein et al., 2007). Unsurprisingly, a large amount of resources is invested in determining how

best to communicate treatment effectiveness of medical intervention (Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, & Woloshin, 2007).

1.7. Making sense of treatment effectiveness

Making sense of treatment effectiveness is not straightforward. Why is it that people do not judge treatment for CMHCs as sufficiently effective to use? Beyond the simple distinction about whether a treatment works or does not work, lies the question of how well a treatment works. To answer this more subtle question a number of different measures are used to describe, communicate, and understand the extent of a treatment's effectiveness. Common measures include the absolute risk reduction (ARR), the number needed to treat (NNT), and the relative risk reduction (RRR).

The ARR is the absolute reduction in rates of adverse outcomes between the therapeutic and the control group. For example, talking-therapy reduces generalised anxiety disorder (GAD) for 32 people in every 100 treated (Hunot et al., 2007). The NNT is an alternative, mostly used among clinicians, and signifies the number of people that need to receive the treatment for one person to benefit from the treatment. For example, 3 people with GAD need to receive talking-therapy for one person to go into remission as a result of treatment (Hunot et al., 2007). It is the multiplicative inverse of the ARR. The RRR, is the proportional reduction of adverse outcomes in the control group. For example, talking therapy reduces GAD by 37% (Hunot et al., 2007). See Table 1.1 for more details on these measures.

Table 1.1: Measures of treatment effectiveness described on the basis of CBT for GAD

	Definition/equation	Example
Outcomes		
Baseline-risk (B_RISK)	Absolute risk of illness in the absence of treatment, (i.e. those who have not recovered by themselves).	0.86 or 86%
Risk in treatment group (RT)	Absolute risk of illness in the treatment group (i.e. those who have not recovered despite treatment).	0.54 or 54%
Measure of treatment effectiveness		
Absolute Risk Reduction (ARR)	ARR = B_RISK – RT	0.32 or 32% (i.e. 0.86 - 0.54 = 0.32)
Relative Risk Reduction (RRR)	RRR = ARR / B_RISK	0.37 or 37% (i.e. 0.32 / 0.86 = 0.37)
Number Needed to Treat (NNT)	NNT = 1/ARR	3 (i.e. 1/0.32 = 3.125)

1.8. Misleading interpretations of treatment effectiveness

Correctly understanding treatment effectiveness is a complex task, even for those with medical training. This is reflected in the observation that the effectiveness of a treatment is perceived differently depending on how the information about its effectiveness is presented (Covey, 2007). For example, the RRR often looks more impressive than the ARR leading to more favourable but less stable judgements of effectiveness (Baron, 1997; Covey, 2007; Gyrd-Hansen, Kristiansen, Nexøe, & Nielsen, 2002; Gyrd-Hansen et al., 2002). This is true particularly at low B_RISKs, at which the RRR looks particularly impressive (Sackett, Richardson, Rosenberg, & Haynes, 1997). The reason is that when most people go into remission in the absence of treatment, even

treating only a small additional number of people is a large reduction in relative terms (i.e. when two people are at risk, a treatment that helps one more person has a 50% RRR). It needs to be remembered that at the beginning of treatment it is not known who will recover naturally, meaning that all participants need to receive the treatment even if 98% would have recovered without it. If at all, it is recommended that the RRR should not be used without presenting the absolute risk in the control group (i.e. B_RISK) to avoid giving an unrealistic impression of the treatment effectiveness (Cook & Sackett, 1995; Gigerenzer, 2003; Schwartz, Woloshin, Dvorin, & Welch, 2006). For the above reasons the ARR is generally preferred (Cook & Sackett, 1995; Gigerenzer, 2003; Schwartz et al., 2006). Understanding the effectiveness of treatments is best when information is presented with the B_RISK, in natural frequencies, with pictographs, and using an incremental risk format (e.g. ARR) to highlight how treatment changes risks from the pre-existing B_RISK (Gigerenzer et al., 2007). The NNT is generally regarded as an appropriate alternative for health professionals (Gigerenzer et al., 2007).

As the health-care service develops, patients are more commonly offered information when making a decision about whether or not to opt for treatment. For example, information about cancer treatments is often presented in terms of the ‘survival rates’, which refers to the proportion of people who are alive after a certain period of time following treatment (e.g. Cancer Research UK, 2013; MayoClinic, 2013); the inverse of the risk remaining in the treatment group. Sophisticated decision tools have been developed to help patients make decisions about whether or not to opt for treatment given a predicted effectiveness (Michaelson, 2011; Wishart et al., 2010). Following guidelines on

how to best present information, these tools present the effectiveness using the B_RISK and the ARR with pictographs. For example, 'Predict' (www.predict.nhs.uk) from the UK National Health Service, for adjuvant therapy following surgery for breast cancer, reports the number of women alive at 5 years with no adjuvant treatment following surgery, the extra number of women treated that are alive because of hormone therapy, and the overall survival rate with adjuvant treatment (Wishart et al., 2010). Similar tools are available for other medical conditions (e.g. J. McCormack, 2013).

1.9. Impact of the B_RISK on perceptions of treatment effectiveness

However, research suggests that even if treatments are presented following current best practice guidelines (Gigerenzer et al., 2007) heuristics may adverse affect decisions about whether treatments are effective (Vogt et al., 2012). Heuristics can be described as thinking short cuts that humans employ to make decision making more efficient but which can lead to errors (Tversky & Kahneman, 1974). Specifically, evidence showed that presenting the B_RISK biases treatment perceptions (Vogt et al., 2012). This finding does not reflect a well know heuristic, the base-rate neglect, when individuals fail to account for the frequency of a characteristic in the population (Bar-Hillel, 1980).

A normative argument would state that the effectiveness of treatments should be judged on nothing but the treatment's effectiveness, the ability to reduce the risk of illness compared to no treatment. For physical health conditions, research demonstrated, however, that people consider treatments to be more effective for conditions in which most people get better by themselves (i.e.

conditions with low B_RISKs), regardless of actual treatment effectiveness (Vogt et al., 2012). In other words, people appear to look at the overall chance of a good outcome when they judge treatment effectiveness (i.e. in effect crediting treatments for good outcomes that result from people getting better by themselves).

To illustrate, imagine there are two groups with 100 patients each. One group has a condition called Lowbaserisk and the other Highbaserisk. They are both anxiety disorders of similar severity. Of those with Lowbaserisk, 99% go into remission after treatment. By contrast, in those with Highbaserisk, 46% go into remission after treatment. But, treatment for Lowbaserisk and Highbaserisk helps the same number of people, namely 32 out of every 100 treated. The large difference in people who continue to suffer is therefore determined by the B_RISK, which is 33% in Lowbaserisk and 86% in Highbaserisk. In other words, people with Lowbaserisk have a much higher rate of getting better by themselves and for that reason most people are in remission after treatment. The research from physical health conditions, described in the preceding paragraph, showed that many people judge a treatment for a condition such as Lowbaserisk as more effective than that for Highbaserisk; even though each treatment helps 32 out of every 100 treated.

Given that people appear to be influenced by the B_RISK when they judge treatments for physical health conditions, the B_RISK may also influence decisions about CMHCs. This is important because in CMHCs, such as generalised anxiety disorder (GAD), most patients (i.e. 86%) remain unwell without treatment (Hunot et al., 2007). This high B_RISK may reduce perceptions that psychological treatments for CMHCs are effective, despite

their substantial impact (i.e. talking-therapy helps 32% of people with GAD). In turn, this may lead to low help-seeking behaviour for CMHCs. This thesis will therefore examine the extent to which beliefs about the effectiveness of talking therapies for CMHCs are influenced by the B_RISK.

1.9.1. Previous research on the B_RISK

Previous studies have described the impact of the B_RISK on treatment and policy decisions. However, these either failed to exclude an alternative explanation, namely differences in treatment effect, or failed to distinguish between treatment effect and B_RISK (e.g. Fetherstonhaugh, Slovic, Johnson, & Friedrich, 1997; Friedrich et al., 1999; Gyrd-Hansen et al., 2002; V. K. Smith & Desvousges, 1987). The challenge in separating the impact of the B_RISK from the treatment effect on treatment and policy decisions is twofold: (i) treatment effects can be measured in different ways, such as the ARR or the RRR, and (ii) they cannot be varied independently, holding constant the B_RISK, an increase in one measure of treatment effect (e.g. ARR) is mirrored with an increase in the other measure of treatment effect (e.g. RRR).

For example, one study asked participants to choose between two treatments of equal ARR. The treatments were ‘Treatment for illness X’ with a B_RISK of 20/1000 and an ARR of 10/1000 and ‘Treatment for illness Y’ with a B_RISK of 80/1000 and an ARR of 10/1000 (Gyrd-Hansen et al., 2002). The resulting RRR for ‘Treatment for illness X’ is 50%, whereas the RRR for ‘Treatment for illness Y’ is only 12.5%. However, the values of the RRR were not actually communicated to participants. The study found that despite the equal ARR, of those willing to choose between treatment X and treatment Y, 59% chose

treatment X whereas 41% opted for treatment Y. The findings were interpreted as *“This result suggests that baseline risk information influences preferences, and that a majority of respondents prefer options that provide a higher RRR (p. 73)”.* While the researchers clearly acknowledge, both, the RRR and the B_RISK they do not distinguish between them, leaving their interpretation ambiguous about whether the preference for treatment X was caused by the lower B_RISK, the higher RRR, or both. Therefore, the study does not provide evidence that the B_RISK influences perceptions about treatments.

In another study, participants were asked to compare the benefit of installing fresh water systems in one of two refugee camps (Fetherstonhaugh et al., 1997). The camps were called Moga 1 and Fizi 1. The fresh water systems would save 1500 people from unclean water in each camp. The camps differed in that prior to the aid, 5% of the water in Moga 1 met current clean water standards and with the aid 50% of the standard would be met. In Fizi 1, prior to the aid, 50% of the water met current clean water standards and with the aid 95% of the standard would be met. Participants were asked about which of the two programs provided more benefit. The findings showed that participants considered the aid program for Fizi 1 to be more beneficial. The findings were interpreted to reflect that the participants had a preference for programs in camps that only needed little extra help. The authors describe this phenomenon as the *“diminished sensitivity in valuing lifesaving interventions against a backdrop of increasing numbers of lives at risk”,* or in short, "psychological numbing"(Fetherstonhaugh et al., 1997, p. 283) or “drop in the bucket effect” (Fetherstonhaugh et al., 1997, p. 285). In this study, prior risk is the B_RISK, and is 95% in Moga 1 and 50% in Fizi 1. Both programs have an

ARR of 45%. However, the researchers do not consider that the RRR for Moga 1 is 47% and the RRR for Fizi 1 is 90%. In other words, the preference for Fizi 1 may have occurred because of the lower B_RISK or the higher RRR; the study therefore does not provide evidence that the B_RISK influences perceptions.

Another study that pursued the “psychological numbing” effect investigated students’ willingness to support mandatory antilock brake requirements for new cars (Friedrich et al., 1999). In the study, participants in two conditions were either informed that the continued use of conventional braking system cost 9000 or 41000 lives per year. In each condition, participants were informed that mandatory antilock brake requirements for new cars would save 150 lives per year. The results showed that participants were more in favour of mandatory antilock brake systems if conventional braking system cost 9000 lives per year. The results were interpreted as participants “value of life-saving interventions in proportional terms” (Friedrich et al., 1999, p. 296). The authors therefore suggest that the observed effect is the result of the impact of the RRR. In the article, the authors do not provide the RRRs, but these can be calculated as 1.6% (i.e. 150/9000) and 0.4% (i.e. 150/41000). Their design and interpretation suggests that the authors do not separate between the B_RISK and the RRR. However, the results of the study can be interpreted as resulting from a preference for a lower B_RISK (i.e. 9000) over a higher B_RISK (i.e. 41000), a preference for a higher RRR (i.e. 1.6%) over a lower RRR (i.e. 0.4%), or both. The study therefore does not provide evidence that the B_RISK, or indeed the RRR, influences preferences because the factors are treated as if they were one and the same.

Yet another study conceptualised the “drop in the bucket” phenomenon as one in which people respond to relative differences (Baron, 1997). In doing so the study also failed to consider that the concepts B_RISK and RRR may have independent effects on decisions. The study asked participants how much they were willing to pay for an extra procedure to be included in their health insurance. Participants were presented with four vignettes: (i) 1000 die of this disease and the procedure will save 900 lives, (ii) 100 die of this disease and the procedure will save 90 lives, (iii) 10000 die of this disease and the procedure will save 900 lives, and (iv) 1000 die of this disease and the procedure will save 90 lives. These vignettes were then construed in a 2 (ARR: 900 vs. 90) x 2 (RRR: 90% vs. 9%) factorial design. The findings showed that participants were willing to pay more for the procedures that saved 900 lives than those that saved 90, and that they were willing to pay more for procedures that reduced the risk by 90% than those that reduced risk by 9%. However, in the study it was not considered that, combined, the procedures that offered the high RRRs (i.e. 90%), also had lower B_RISKs (i.e. 100 and 1000) compared to the procedures that offered the low relative reduction (i.e. 9%), which had higher B_RISKs (i.e. 1000 and 10000). It is thus possible that the greater willingness to pay for procedures that reduced the risk by 90% was caused by the lower B_RISKS (i.e. 100 and 1000 vs. 1000 and 10000). The study therefore failed to control for the interconnectivity of the ARR, the RRR, and the B_RISKs.

While the treatment effect and the B_RISK cannot be varied independently, their interconnectivity is non-parallel: an increase in the ARR increases the RRR more strongly at a lower B_RISK than at a higher B_RISK. This non-

parallel interconnectivity can be exploited and was used to establish the independent impact of the B_RISK on how effective treatments for physical health conditions are perceived to be, while controlling for treatment effects as measured by the ARR and the RRR (Vogt et al., 2012). The current study will also exploit this non-parallel interconnectivity.

1.10. Impact of the ARR and the RRR on perceptions of treatment effectiveness

While it was established that the B_RISK had an independent impact on perceptions about treatment effectiveness (Vogt et al., 2012), previous studies were not designed to examine whether the ARR and/or the RRR also had an independent impact on treatment perceptions. The current study will therefore also examine whether the ARR and/or the RRR have an independent impact on the perceived effectiveness of talking-therapy when the B_RISK is presented.

1.11. Moderators

1.11.1. Mood

Decisions about the effectiveness of talking-therapy for CMHCs may also be influenced by the mood of the person making the decision. People often respond in coherence with their mood (de Vries, Holland, Corneille, Rondeel, & Witteman, 2012). This suggests that people with low mood may consider talking-therapy for CMHCs more negatively (i.e. less effective). However, people with low mood also pay more attention to the rules that determine utilitarian decisions and are less influenced by experience (de Vries et al.,

2012). This suggests that while people with low mood may judge the effectiveness of talking-therapy as less effective, they may, at the same time be less influenced by the B_RISK. The effects of mood will be controlled for in the analyses to minimise a confounding effect.

1.11.2. Numeracy

The search for the optimal way of communicating treatment information is driven by the acknowledgement that medical information is often complex and that people can find it difficult to understand this information (Gigerenzer et al., 2007). There is also an understanding that many people have poor numeracy skills, for example difficulties with a broad range of concepts, including fractions, proportions, and probability judgments (Reyna & Brainerd, 2007). A large body of evidence shows that poor numeracy skills predict poorer health outcome, less accurate perceptions of health risks, and a compromised ability to make medical decisions (Ancker, Senathirajah, Kukafka, & Starren, 2006; Brown et al., 2011; Galesic, Garcia-Retamero, & Gigerenzer, 2009; Reyna & Brainerd, 2007).

Despite the general consensus that poor numeracy is associated with poorer medical decisions, previous research examining the B_RISK did not detect that numeracy moderated the impact of the B_RISK on the perceived effectiveness of treatments for physical conditions (Vogt et al., 2012). In other words, the B_RISK seemed to influence people's perceptions of treatment effectiveness independent of whether they had lower or higher numeracy skills. However, numeracy is commonly seen to affect decision making, and therefore the

effects of numeracy will be controlled for in the analyses to minimise a confounding effect.

1.12. Research aims and hypotheses

The aim of this research is to examine the possible impact of the B_RISK and the actual treatment effectiveness, as measured by the ARR and the RRR, on perceptions of the effectiveness of talking-therapy. The importance of this question is based on the fact that perceptions of effectiveness are a key predictor of people using talking-therapy for CMHCs. The hypotheses are:

Hypothesis 1: Talking-therapies for conditions with higher B_RISKs are perceived as less effective.

Hypothesis 2: Talking therapies with higher ARRs are perceived as more effective.

Hypothesis 3: Talking therapies with higher RRRs are perceived as more effective.

2. Method

2.1. Design

The analysis is designed as three 2 x 2 x 2 x 4 mixed within-between subjects factorial experiments:

For Hypothesis 1, the within-subjects factors are the ARR (two levels: low vs. high) and the RRR (two levels: low vs. high), and the between subjects factors are mood (two levels: low vs. high) and numeracy (four levels: poor vs. fair vs. good vs. excellent).

For Hypothesis 2, the within-subjects factors are the B_RISK (two levels: low vs. high) and the RRR (two levels: low vs. high), and the between subjects factors are mood (two levels: low vs. high) and numeracy (four levels: poor vs. fair vs. good vs. excellent).

For Hypothesis 3, the within-subjects factors are the B_RISK (two levels: low vs. high) and the ARR (two levels: low vs. high), and the between subjects factors are mood (two levels: low vs. high) and numeracy (four levels: poor vs. fair vs. good vs. excellent).

*2.2. **Procedure***

The study was conducted online. Participants were recruited through an advertisement (Appendix 1) placed via Google Ltd in the online advertisement spaces on the main Google websites as well as partnering organisations. Advertisements were directed to those above 18 years of age, limited to UK, and limited to tablets and computer access. Advertisement was displayed in

response to key words being used that related to CMHCs (Appendix 2). Those who clicked on the advertisement banner were directed to the study's webpage. On the website participants were first provided with an information sheet describing the study. Participants could agree to participate by consenting to the study and were then directed to the questionnaire (Appendix 4). In total, 1244 people clicked on the advertisement, of which 365 consented and arrived at the first page of the questionnaire between July 2013 and January 2014. Completion of items relating to the dependent variable (i.e. perceived effectiveness of talking-therapy) was not optional. Thus, if participants did not complete these items they could not progress to the next page. Of the 365 people consenting to the study, 210 completed the items relating to the dependent variable. Of the 210 people completing the questionnaire, 6 identified themselves as below 18 years of age and were excluded from the analysis. Two participants entered implausible data and were also excluded (more information about these two cases is provided in section 2.8).

In the questionnaire, all participants were shown six vignettes depicting the effectiveness of talking-therapy for a CMHC. Each of the six vignettes was followed by a question asking participants about how effective they perceive talking-therapy was for the depicted CMHC (i.e. the dependent variable). All six vignettes were shown on one page, vertically, so that participants could scroll down the page to access all vignettes. Once participants had completed this part of the questionnaire, they were directed to the next section. There, participants provided information about their demographic details, their mood, experience with health problems, and numeracy. Finally, participants were

debriefed and offered accurate information about the effectiveness of talking-therapies for CMHCs based on current evidence from the Cochrane library or the NICE.

The study received ethical approval from the Royal Holloway Psychology Department Ethics Committee (Appendix 3). Pilot work demonstrated that completion of the entire study took approximately 10 minutes. Additional measures were collected but these are not part of this thesis (see Appendix 4).

2.3. Participants

Participants comprised a sample of the adult population (n=202). The mean age was 47.5 years with a standard deviation of 13.8 years. Age ranged from 18 to 80 with good representation across all ages (Figure 2.1). Of the sample, 26% were men, 93.5% were white. Of the participants, 55.3% had achieved at least A' level or similar, or admission to university. At least two of the numeracy questions were correctly answered by 39.4% of participants. The majority, 61.3%, had received treatment for a mental health problem in the past, and 42.2% were currently receiving treatment for a mental health problem. For 156 participants the location of where the questionnaire had been completed was collected automatically by the internet server. This automatically collected data suggested that participants came from all across the UK. Data from 105 different cities was detected, with most participants coming from London (see Figure 2.2). See Table 2.1 for participants' characteristics.

Figure 2.1: Distribution of participant ages

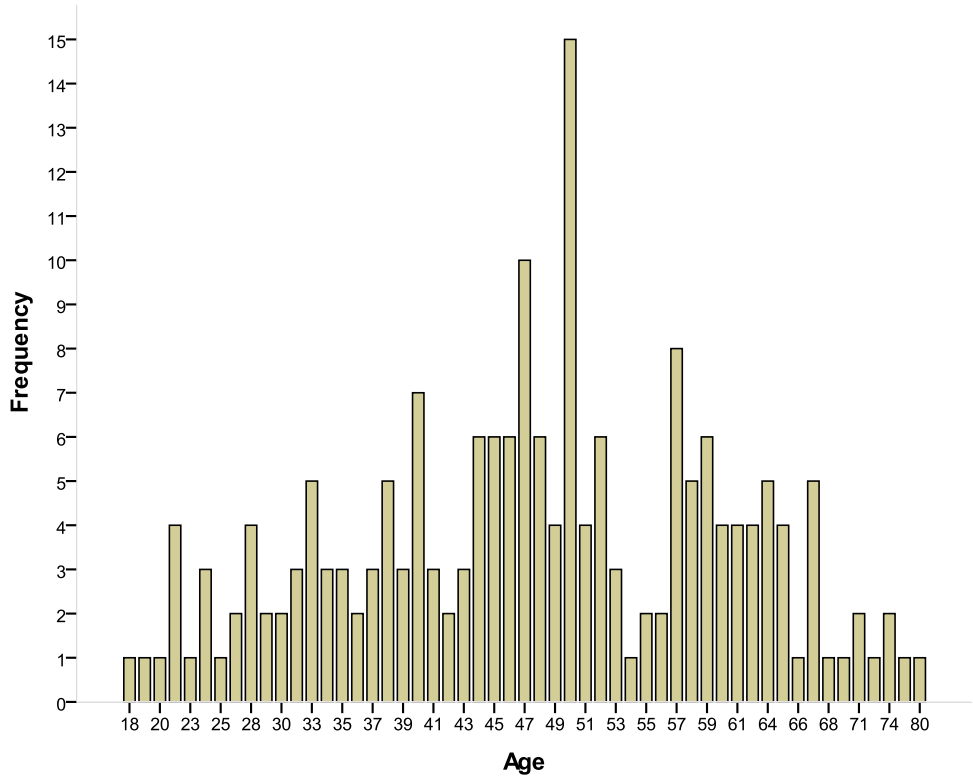


Figure 2.2: Location of participants in the UK

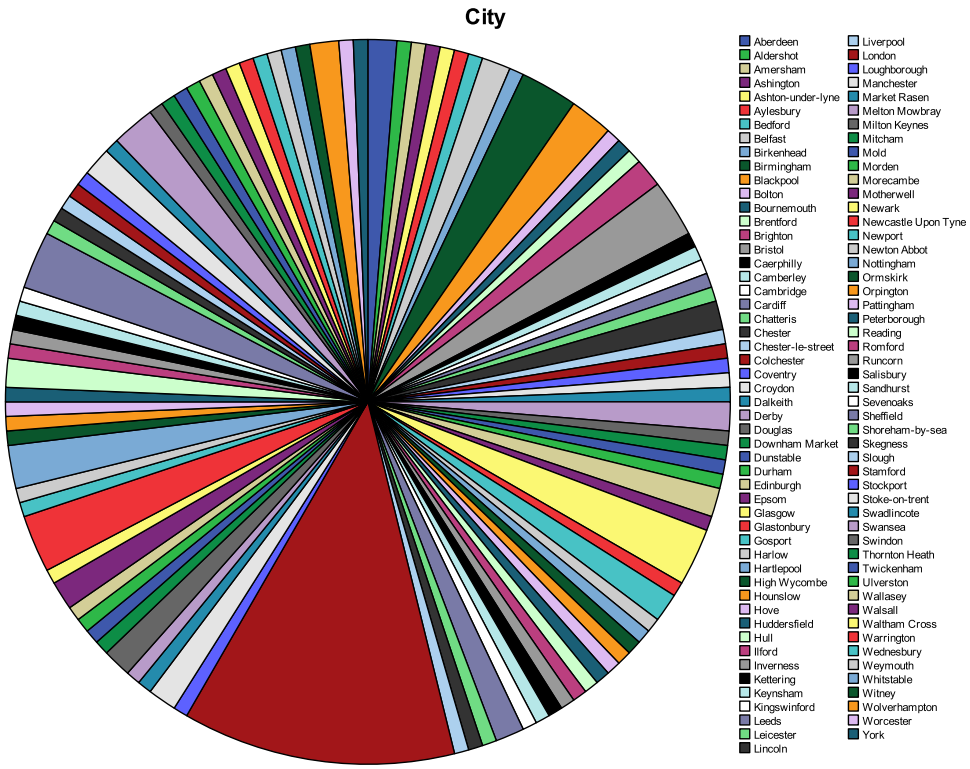


Table 2.1 Participant characteristics (N = 202)

Age (n=190)	
Mean (Std. Deviation)	47.65 (13.09)
Minimum, Maximum	18, 80
Gender (n=195)	
Male(n)	26.15% (51)
Female (n)	73.84% (144)
Ethnicity, (n=199)	
White (n)	93.97% (187)
Mixed (n)	1.51% (3)
Black or Black British (n)	0.50% (1)
Asian or Asian British (n)	4.02% (8)
Chinese or other ethnic group (n)	0.0% (0)
Education (n=199)	
No education completed	4.52% (9)
Primary (primary school)	3.02% (6)
Secondary 1 (compulsory education, GCSE, O level, or similar)	37.19% (74)
Secondary 2 (admission to university, A level or similar)	14.57% (29)
Tertiary (university and other after the secondary level)	40.70% (81)
Numeracy (n=202)	
0 correct answers	24.75% (50)
1 correct answer	35.64% (72)
2 correct answers	25.24% (51)
3 correct answers	14.36% (29)
Previous treatment for mental health problem (n=199)	
Yes (n)	61.11% (121)
No (n)	38.89% (77)
Current treatment for mental health problem (n=199)	
Yes (n)	41.91% (83)
No (n)	58.08% (115)
PHQ2 (n=193)	
Mean (Std. Deviation)	4.07 (1.99)
≥3	73.6% (142)
GAD2 (n=191)	
Mean (Std. Deviation)	4.18 (1.93)
≥3	76.4% (146)
PHQ4 (n=186)	
Mean (Std. Deviation)	8.34 (3.56)
≥6	75.8% (141)

2.4. Measures

2.4.1. Perceived effectiveness of treatment

The perceived effectiveness of talking-therapy presented by a vignette was assessed with one item, displayed beneath each vignette. The item read: “How effective is talking-therapy for this condition compared to no treatment?” The item was measured, using a continuous visual analogue scale, response

options ranging from ‘ineffective’ (0) to ‘extremely effective’ (100). To respond, participants could drag a visual pointer between the two endpoints, indicating their judgement. Visual analogue scales are comparable in sensitivity and reproducibility to Likert scales (H. M. McCormack, Horne, & Sheather, 1988) although appear to be slightly superior (Grant et al., 1999). Visual analogue scales have also found to be successful at evaluating the effectiveness of treatments and found to be superior to category rating scales (Carling et al., 2008).

2.4.2. Mood

As mood may be a confounding factor in participants’ responses (i.e. people who are experiencing symptoms of depression or anxiety may respond differently from those who are not) it was necessary to control for mood. Mood was measured using the Patient Health Questionnaire-4 (PHQ-4) (Kroenke, Spitzer, Williams, & Löwe, 2009). The PHQ-4 is an ultra-brief self-report questionnaire based on commonly used measures of depression and anxiety (PHQ-9, Kroenke, Spitzer, & Williams, 2001; GAD-7, Spitzer, Kroenke, Williams, & Lowe, 2006).

The PHQ-4 consists of a 2-item depression scale (PHQ-2) (Kroenke, Spitzer, & Williams, 2003) and a 2-item anxiety scale (GAD-2) (Kroenke, Spitzer, Williams, Monahan, & Lowe, 2007). The PHQ-2 has been found to be a valid screening tool for major depression in adults and older adults (Kroenke et al., 2003; Li, Friedman, Conwell, & Fiscella, 2007; Löwe, Kroenke, & Gräfe, 2005). In a review of ultrashort measures of anxiety the GAD-2 instrument performed best, and was the one measure that met two key criteria with good diagnostic

accuracy and feasibility (NICE, 2011a). It is recommended as an identification tool for anxiety disorders in primary care, in particular GAD, by NICE (NICE, 2011a). The combined PHQ-4 has also been shown to be reliable and valid in a large general population sample (Löwe et al., 2010).

Whereas the PHQ-2 consists of the two DSM-IV diagnostic core criteria for depressive disorders (American Psychiatric Association, 2000), the GAD-2 consists of the two core criteria for GAD, which have also been shown to be effective screening items for panic, social anxiety, and PTSD (Kroenke et al., 2007). Equivalent to the parent scales, the PHQ-4 begins with the stem question: “Over the last 2 weeks, how often have you been bothered by the following problems?”. The two PHQ-2 items are: “Little interest or pleasure in doing things” and “Feeling down, depressed, or hopeless”. The two GAD-2 items are: “Feeling nervous, anxious or on edge” and “Not being able to stop or control worrying”. Response options are ‘not at all’, ‘several days’, ‘more than half the days’, and ‘nearly every day’, scored as 0, 1, 2, and 3, respectively.

The four items were combined into a composite four-item scale. Therefore, the total score ranges from 0 to 12. For the PHQ-2 and the GAD-2, scale scores of ≥3 are suggested as cut-off points between the normal range and probable cases of depression or anxiety, respectively (Kroenke et al., 2007, 2003; Löwe et al., 2010). For the PHQ-4 a score of ≥6 has been described as indicating the possible presence of a depressive or an anxiety disorder (Löwe et al., 2010). The PHQ-4 scores were then categorised into two groups, those scoring below the clinical cut-off (i.e. <6) and above cut-off (i.e. ≥ 6). Splitting the data in this way achieved a balance between making the variable useable and interpretable in an ANCOVA and keeping it clinically relevant.

2.4.3. Numeracy

As stated, previous research has suggested that numeracy skills are related to medical decision making. In order to control for numeracy, it was measured using three standard items forming a global numeracy scale that is not specific to health numeracy (Schwartz, Woloshin, Black, & Welch, 1997). When compared to specific health numeracy measures, this 3-item scale was found to measure the same underlying global numeracy construct (Lipkus, Samsa, & Rimer, 2001).This non-health specific short measure is now common among studies investigating numeracy and the communication of medical risks (e.g. Estrada, Barnes, Collins, & Byrd, 1999; Galesic et al., 2009; Vogt et al., 2012).

The three items were: (i) “Imagine that we flip a fair coin 1,000 times. What is your best guess about how many times the coin would come up heads in 1,000 flips?”; (ii) “In the Big Cash Lottery, the chance of winning a £10 prize is 1%. What is your best guess about how many people would win a £10 prize if 1000 people each buy a single ticket for Big Cash Lottery?”; (iii) “In the Bargain Cars Sweepstakes, the chance of winning a car is 1 in 1,000. What percent of tickets from Bargain Cars Sweepstakes win a car?”. Participants could score three points if they answered all three questions correctly. Responses were combined into a 4-point scale (i.e. 0 = poor numeracy, 1 = fair numeracy, 2 = good numeracy, 3 = excellent numeracy).

2.4.4. Demographic information

To provide information about the participants’ characteristics, demographic information was collected about age, gender, ethnicity, and education. These variables are also used as covariates in the analysis.

Ethnicity

Standard procedures were used for measuring ethnicity in the UK following the 2001 census presentation (Statistics, 2003). The item read: 'What is your ethnic group?'. The response groups included: White, mixed, Asian or Asian British, Black or Black British, Chinese or other ethnic group. For the analysis the groups were collapsed into 'white' vs. 'non-white'.

Education

Education was measured using one item, "What is your highest level of education completed?" (Vogt et al., 2012). The response options included: No education completed, Primary (primary school), Secondary 1 (compulsory education, GCSE, O level, or similar), Secondary 2 (admission to university, A level or similar), Tertiary (university and other forms of education after the secondary level).

Other mental health information

Information about current and past mental health treatment was collected. As well as providing information about the sample of participants, these variables were used as covariates. To assess this, measures of (i) current receipt of treatment for the mental health disorder, and (ii) previous receipt of treatment for the mental health disorder, were included. The items read: "Are you currently receiving pharmacological or talking-therapy treatment for a mental health problem?", and "Have you previously received pharmacological or talking-therapy treatment for a mental health problem?". The response options for both measures were 'yes' or 'no'.

2.5. Vignettes

Six vignettes were used to depict the effectiveness of talking-therapy for a CMHC. Participants were asked to imagine that they had been diagnosed with a condition. The six CMHCs were (i) chronic low self-esteem, (ii) depression, (iii) OCD, (iv) GAD, (v) PTSD, and (vi) social anxiety disorder. Each CMHC was briefly described in a pop-up window (Table 2.2).

Six different outcomes were presented to participants (Table 2.3). These outcomes were hypothetical and this was described to participants before they saw the vignettes. The instructions read *"The information shown about the effectiveness of talking-therapy on this page is fictional. Imagine that you have been diagnosed with* [one of six CMHCs]*. Treatment includes talking-therapy. In the graphic below you can see the outcomes for people who choose no treatment and for people who choose talking-therapy"*.

Table 2.2: Brief pop-up window descriptors of conditions

Chronic low self-esteem	"Self-esteem refers to the overall opinion we have of ourselves and the value we place on ourselves as people. Low self-esteem means that the tone of this opinion is negative: for example, 'I'm unlovable' or 'I'm useless'. For some people low self-esteem can persist for a long time, that is, chronic"
Depression	"Depression is more than simply feeling unhappy or fed up for a few days. We all go through spells of feeling down, but when you're depressed you feel persistently sad for weeks or months, rather than just a few days. Depression affects people in different ways and can cause a wide variety of symptoms. They range from lasting feelings of sadness and hopelessness, to losing interest in the things you used to enjoy and feeling very tearful. Many people with depression also have symptoms of anxiety. There can be physical symptoms too, such as feeling constantly tired, sleeping badly, having no appetite or sex drive, and complaining of various aches and pains."
Obsessive compulsive disorder (OCD)	"Obsessive Compulsive Disorder (OCD) is a mental health condition where a person has obsessive thoughts and compulsive behaviour. For example, someone who is obsessively scared they will catch a disease may feel the need to have a shower every time they use a toilet."
Generalised anxiety disorder (GAD)	"People with Generalised Anxiety Disorder (GAD) find it hard to control their worries. Their feelings of anxiety are more constant and often affect their daily life. GAD is a long-term condition which causes you to feel anxious about a wide range of situations and issues, rather than one specific event. People with GAD feel anxious most days and often struggle to remember the last time they felt relaxed."
Post-traumatic stress disorder (PTSD)	"Post-Traumatic Stress Disorder (PTSD) is an anxiety disorder caused by very stressful, frightening or distressing events. Someone with PTSD will often relive the traumatic event through nightmares and flashbacks, and they may experience feelings of isolation, irritability and guilt. They may also have problems sleeping, such as insomnia, and find concentrating difficult. These symptoms are often severe and persistent enough to have a significant impact on the person's day-to-day life."
Social anxiety disorder	"Social Anxiety Disorder is an anxiety disorder characterized by intense fear in social situations causing considerable distress and impaired ability to function in at least some parts of daily life. Social anxiety disorder typically involves a persistent, intense, chronic fear of being judged by others and of being humiliated by one's own actions. Physical symptoms often accompanying social anxiety disorder include excessive blushing, sweating, trembling, palpitations and nausea."

Table 2.3: Combinations of treatment outcomes

Label	ARR	B_RISK	RRR
Vignette 5/10=50	5%	10%	50%
Vignette 15/30=50	15%	30%	50%
Vignette 5/30=17	5%	30%	17%
Vignette 15/90=17	15%	90%	17%
Vignette 5/90=6	5%	90%	6%
Vignette 45/90=50	45%	90%	50%

The outcomes were presented with pictographs using icon-arrays and with written text. A screenshot of how the vignettes were presented is shown in Figure 2.3. The information depicting the effectiveness of talking therapy for the respective CMHCs was designed according to current best practice guidelines for communicating medical risks and treatments (Fagerlin, Zikmund-Fisher, & Ubel, 2011; Zikmund-Fisher, Fagerlin, & Ubel, 2010). Understanding the effectiveness of treatments is best when information is presented with the B_RISK, in natural frequencies, with pictographs, and using an incremental risk format (i.e. ARR) to highlight how treatment changes risks from the pre-existing B_RISK (Fagerlin et al., 2011; Gigerenzer et al., 2007).

The icon arrays showed the outcomes for 100 people with the respective condition who did not receive treatment after 5 years: (i) symptom free with blue dots, and (ii) not symptom free with red dots (i.e. B_RISK). The outcomes for 100 people with the respective condition who received treatment after 5 years were depicted as: (i) symptom free with blue dots, (ii) symptom free because of talking-therapy with green dots (i.e. ARR), and (iii) not symptom free with red dots (i.e. risk in treatment group).

Figure 2.3: Screenshot of outcomes for conditions

The written text described those with the respective condition who did not receive treatment who were symptom free after 5 years *"[xx] out of 100 people would be symptom free in 5 years"*. Written text also described those with the respective condition who received treatment and had been helped by the treatment (i.e. ARR) *"[xx] more people out of 100 would be symptom free because of talking-therapy in 5 years"*.

A potential contaminating factor is that participants may have specific views about certain CMHCs. For example, they may view recovery without treatment as more likely from one CMHC than another. To avoid specific view about certain CMHCs affecting the results, the six CMHCs and the six outcomes were arranged differently, at random, for each participant. This was achieved with an algorithm computed in the online questionnaire. The order in which the vignettes were shown was also randomised. A similar method has been used previously (Vogt et al., 2012).

The challenge in testing whether the B_RISK, the ARR, and the RRR have independent impacts on the perceived treatment effectiveness, is that they are linked to each other and cannot be varied independently. For example, holding constant the B_RISK (e.g. at 30%), an increase in one measure of treatment effectiveness (e.g. ARR: 5 to 15) is mirrored with an increase in the other measure of treatment effectiveness (e.g. RRR: 17% to 50%). However, while treatment effect and B_RISK cannot be varied independently, their relation is not parallel. This non-parallel relation can be used to assess the impact of each factor on perceptions of treatment effectiveness. The reason is that perceptions of treatment effectiveness should follow this non-parallel relationship between the ARR, the RRR, and the B_RISK, if each factor influences perceptions of treatment effectiveness. The outcomes on the six vignettes were selected specifically to test the non-parallel relationships. They were thus combined to test the independent impacts of the ARR, the RRR, and the B_RISK.

2.5.1. Vignettes for Hypothesis 1: B_RISK

An increase in the ARR (i.e. 5 to 15), at a low RRR (i.e. RRR of 17%) increases the B_RISK strongly (i.e. B_RISK: 30% to 90%), whereas the same increase in the ARR at a large RRR (i.e. RRR of 50%), increases the B_RISK less (i.e. B_RISK: 10% to 30%; see Table 2.4), manifesting in a non-parallel relation between the three variables (see Figure 2.4).

This non-parallel relation can be used to assess the postulated negative impact of the B_RISK on perceptions of treatment effectiveness, because perceptions of treatment effectiveness should follow this non-parallel relationship if the B_RISK influences treatment perceptions. Thus, if higher B_RISKs negatively influence treatment perception then the rapid increase in the B_RISK seen at a RRR of 17% and a ARR of 15% should result in suppressed perceptions of treatment effectiveness in Vignette 15/90=17.

Figure 2.4: Non-parallel relation holding constant the ARR and the RRR

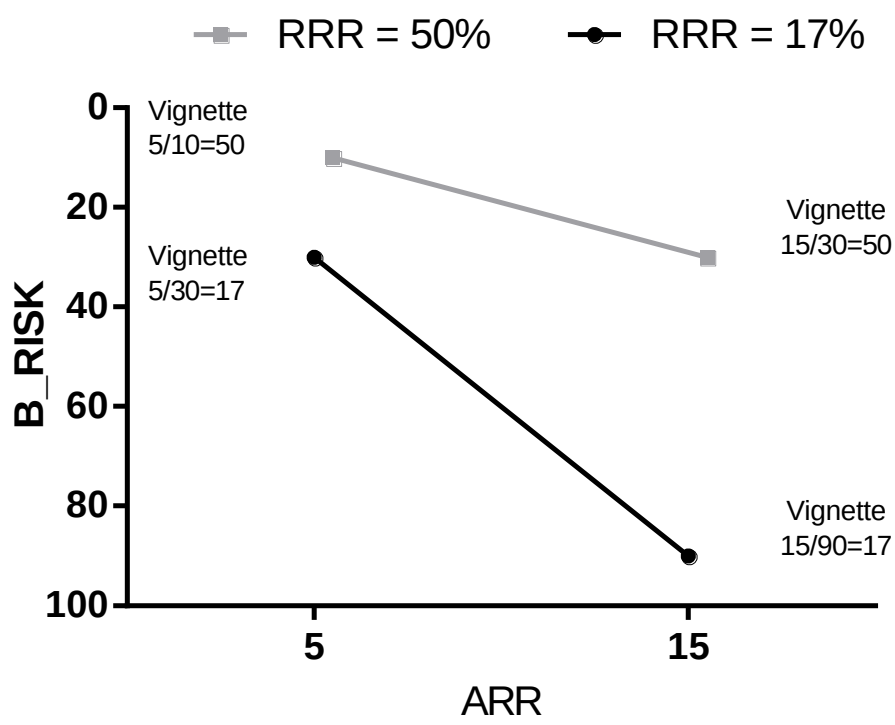
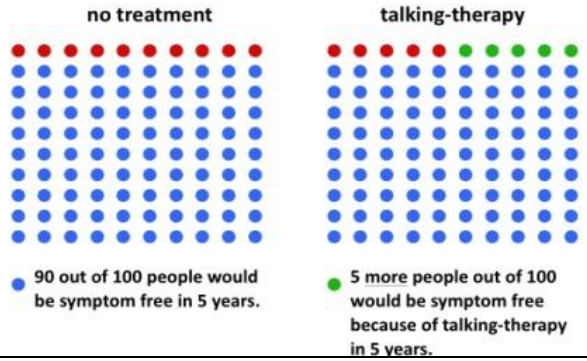
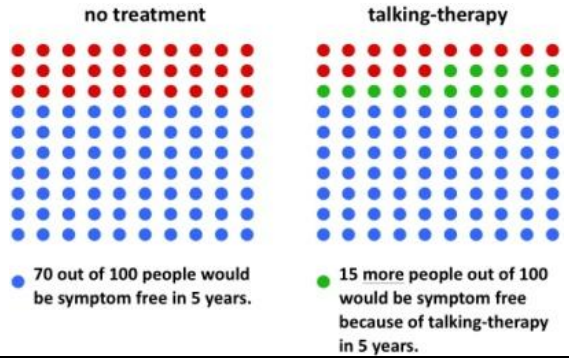
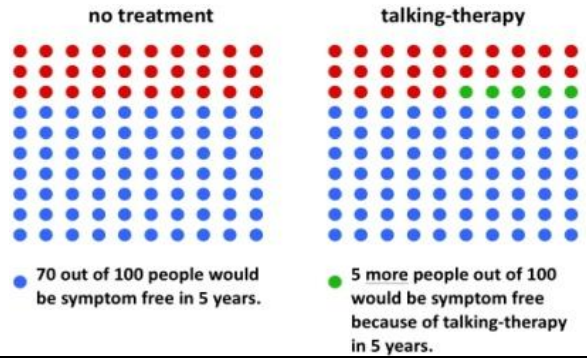
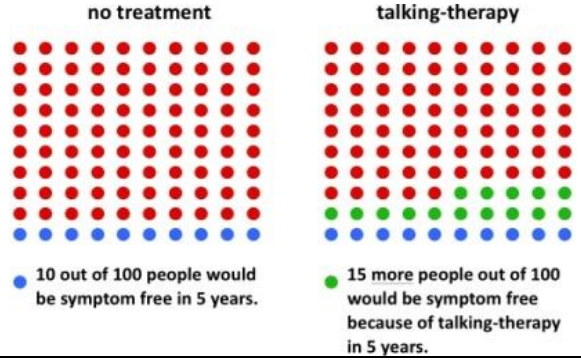


Table 2.4: Vignettes used to keeping the differences in the ARR and the RRR constant

	ARR = 5%	ARR = 15%	Difference
RRR=50%	**Vignette 5/10=50** ARR = 5% B-RISK = 10% RRR = 50% 	**Vignette 15/30=50** ARR = 15% B-RISK = 30% RRR = 50% 	ARR_diff = 10% B-RISK_diff = 20% RRR_diff = 0%
RRR=17%	**Vignette 5/30=17** ARR = 5% B-RISK = 30% RRR = 17% 	**Vignette 15/90=17** ARR = 15% B-RISK = 90% RRR = 17% 	ARR_diff = 10% B-RISK_diff = 60% RRR_diff = 0%
Difference	ARR_diff = 0% B-RISK_diff = 20% RRR_diff = 33%	ARR_diff = 0% B-RISK_diff = 60% RRR_diff = 33%	Total ARR_diff = 0% Total B-RISK_diff = 40% Total RRR_diff = 0%

2.5.2. Vignettes for Hypothesis 2: ARR

An increase in the B_RISK (i.e. 30% to 90%), at a low RRR (i.e. 17%) increases the ARR less (i.e. 5% to 15%), whereas the same increase in the B_RISK at a large RRR (i.e. 50%), increases the ARR more (i.e. 15% to 45%; see Table 2.5), manifesting in a non-parallel relation between the three variables (see Figure 2.5).

This non-parallel relation can be used to assess the impact of the ARR on perceptions of treatment effectiveness, because perceptions of treatment effectiveness should follow this non-parallel relationship if the ARR influences treatment perceptions. Thus, if a higher ARR positively influences treatment perception then the rapid increase in the ARR seen at a RRR of 50% and a B_RISK of 90% should boost perceptions of treatment effectiveness in Vignette 45/90=50.

Figure 2.5: Non-parallel relation holding constant the B_RISK and the RRR

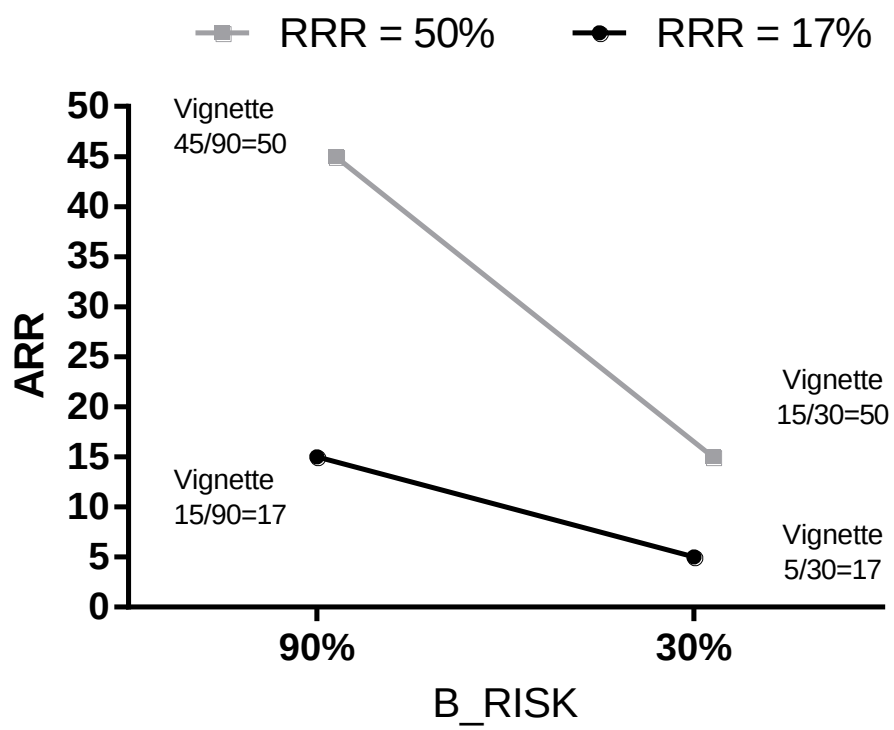
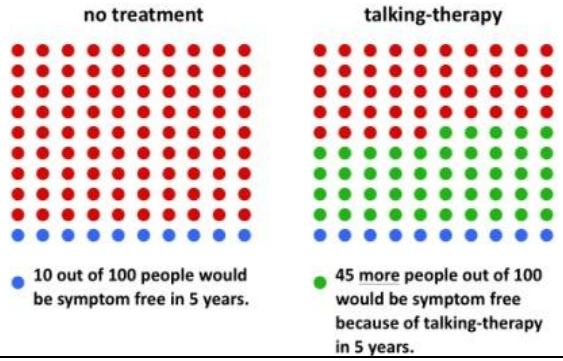
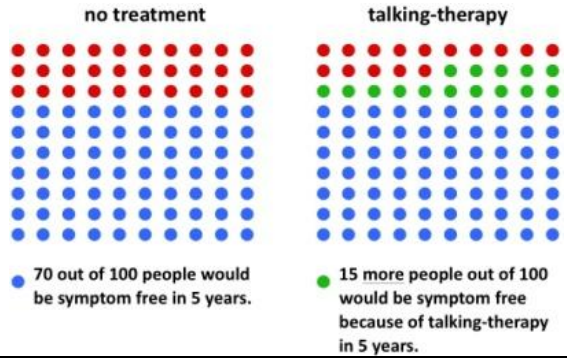
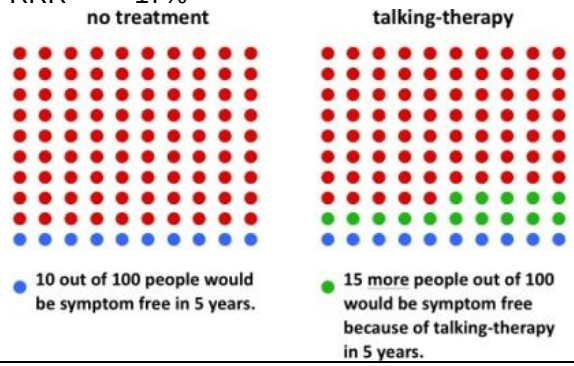
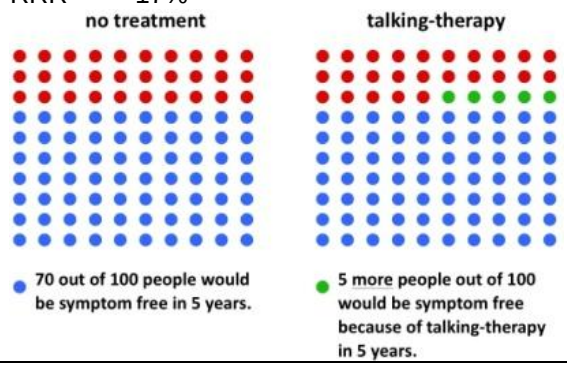


Table 2.5: Vignettes used to keeping the differences in the B_RISK and the RRR constant

	B-RISK = 90%	B-RISK = 30%	Difference
RRR=50%	**Vignette 45/90=50** ARR = 45% B-RISK = 90% RRR = 50% 	**Vignette 15/30=50** ARR = 15% B-RISK = 30% RRR = 50% 	ARR_diff = 30% B-RISK_diff = 60% RRR_diff = 0%
RRR=17%	**Vignette 15/90=17** ARR = 15% B-RISK = 90% RRR = 17% 	**Vignette 5/30=17** ARR = 5% B-RISK = 30% RRR = 17% 	ARR_diff = 10% B-RISK_diff = 60% RRR_diff = 0%
Difference	ARR_diff = 30% B-RISK_diff = 0% RRR_diff = 33%	ARR_diff = 10% B-RISK_diff = 0% RRR_diff = 33%	Total ARR_diff = 20% Total B-RISK_diff = 0% Total RRR_diff = 0%

2.5.3. Vignettes for Hypothesis 3: RRR

An increase in the ARR (i.e. 5% to 15%), at a high B_RISK (i.e. 90%) increases the RRR less (i.e. 6% to 17%), whereas the same increase in the ARR at a low B_RISK (i.e. 30%), increases the RRR more (i.e. 17% to 50%; see Table 2.6), manifesting in a non-parallel relation between the three variable (see Figure 2.6).

This non-parallel relation can be used to assess the impact of the RRR on perceptions of treatment effectiveness, because perceptions of treatment effectiveness should follow this non-parallel relationship if the RRR influences treatment perceptions. If the RRR influences treatment perception then the rapid increase in the RRR seen at a B_RISK of 30% and an ARR of 15% should boost perceptions of treatment effectiveness in Vignette 15/30=50.

Figure 2.6: Non-parallel relation holding constant the ARR and the B_RISK

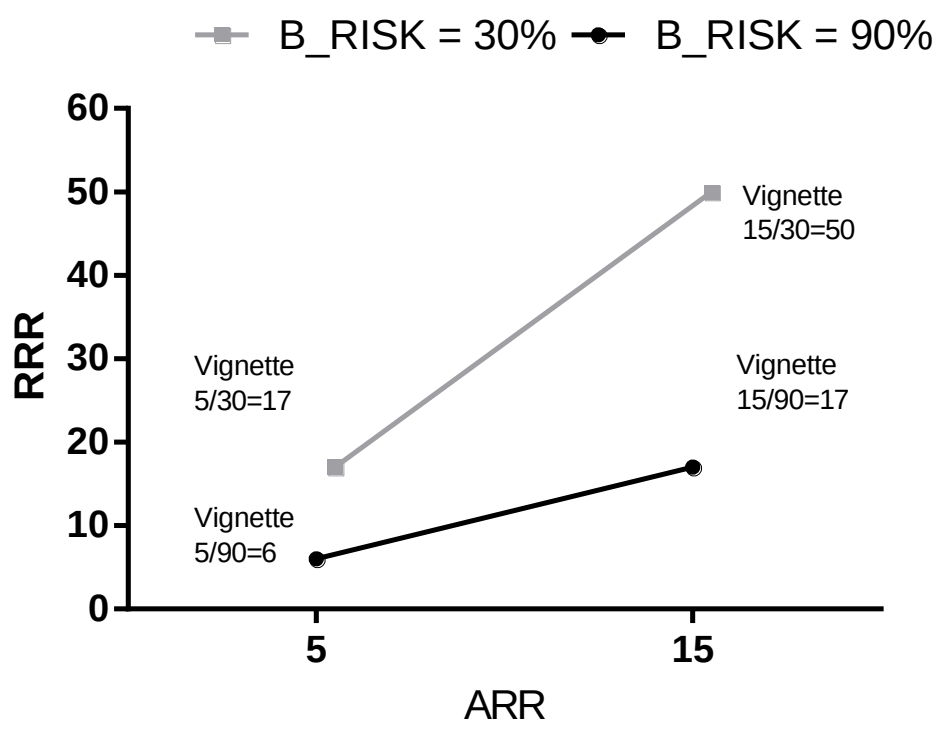
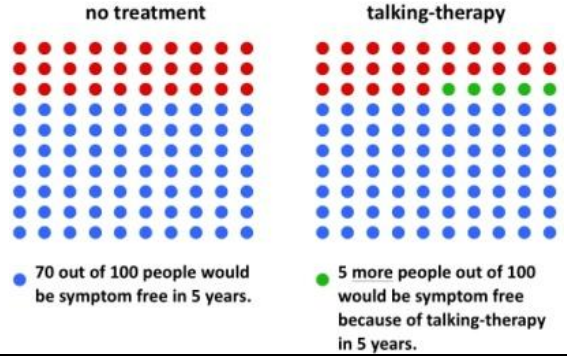
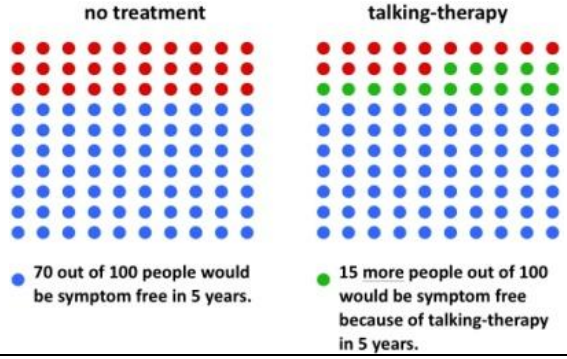
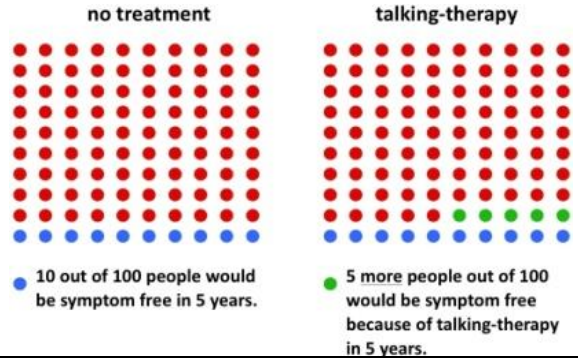
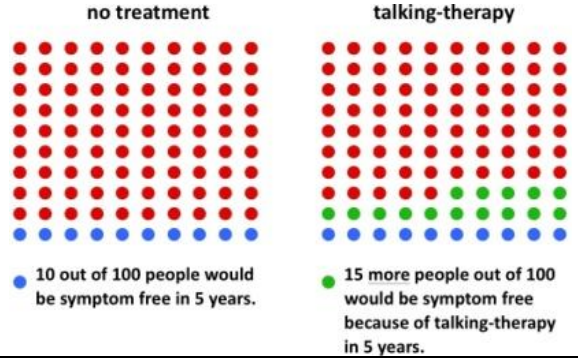


Table 2.6: Vignettes used to keeping the differences in the ARR and the B_RISK constant

	ARR = 5%	ARR = 15%	Difference
B-RISK = 30%	**Vignette 5/30=17** ARR = 5% B-RISK = 30% RRR = 17% 	**Vignette 15/30=50** ARR = 15% B-RISK = 30% RRR = 50% 	ARR_diff = 10% B-RISK_diff = 0% RRR_diff = 33%
B-RISK = 90%	**Vignette 5/90=6** ARR = 5% B-RISK = 90% RRR = 6% 	**Vignette 15/90=17** ARR = 15% B-RISK = 90% RRR = 17% 	ARR_diff = 10% B-RISK_diff = 0% RRR_diff = 11%
Difference	ARR_diff = 0% B-RISK_diff = 60% RRR_diff = 11%	ARR_diff = 0% B-RISK_diff = 60% RRR_diff = 33%	Total ARR_diff = 0% Total B-RISK_diff = 0% Total RRR_diff = 22%

2.6. Analyses

The analysis will commence by using simple bi-variate correlations across the variables. For the main analyses three separate 2 x 2 x 2 x 4 mixed within-between factorials were constructed with the perceived effectiveness of talking-therapy as the dependent variable, as outlined below. All analyses were conducted with IBM SPSS Version 19.

2.6.1. Hypothesis 1: 'Talking-therapies for conditions with higher B_RISKs are perceived as less effective.'

In order to test Hypothesis 1, a 2 x 2 x 2 x 4 mixed within-between subjects ANCOVA was used. The two within subjects factors are ARR (two levels: 5 vs. 15) and RRR (two levels: 17% vs. 50%). The two between-subjects factors were mood (two levels: low vs. high) and numeracy (four levels: poor vs. fair vs. good vs. excellent). If a higher B_RISK negatively influences treatment perception then this would be indicated by an interaction between the RRR and the ARR: perceptions of treatment effectiveness should be suppressed by the large B_RISK (i.e. 90%) at a small RRR (i.e. 17%) and large ARR (i.e. 15%). If mood or numeracy moderated the impact of the B_RISK, then a three-way interaction would be seen between the ARR, the RRR, and mood or numeracy, respectively. ANCOVA is different from ANOVA because it allows the inclusion of covariates in order to control for potential confounding effects which can make the analysis statistically more efficient as compared to including these variables as additional factors in an ANOVA. The covariates can be continuous or dichotomous variables (Tabachnick & Fidell, 2001). A number of covariates

were added, including age, gender, education, and previous and past treatment.

2.6.2. Hypothesis 2: ‘Talking-therapies with higher ARRs are perceived as more effective.’

In order to test Hypothesis 2, a 2 x 2 x 2 x 4 mixed within-between subjects ANCOVA was used. The two within subjects factors are B_RISK (two levels: 30% vs. 90%) and RRR (two levels: 17% vs. 50%). The two between-subjects factors are mood (two levels: low mood vs. high mood) and numeracy (four levels: poor vs. fair vs. good vs. excellent). If a higher ARR positively influences treatment perceptions then this would be indicated by an interaction between the RRR and the B_RISK: perceptions of treatment effectiveness would be boosted by the large ARR (i.e. 45%) at a large B_RISK (i.e. 90%) and a large RRR (i.e. 50%). If mood or numeracy moderated the impact of the ARR, then a three-way interaction would be expected between the B_RISK, the RRR and mood or numeracy, respectively. A number of covariates were included to control for confounding factors, including age, gender, education, and previous and past treatment.

2.6.3. Hypothesis 3 ‘Talking-therapies with higher RRRs are perceived as more effective’

In order to test Hypothesis 3, a 2 x 2 x 2 x 4 mixed within-between subjects ANCOVA was used. The two within subjects factors are B_RISK (two levels: 30% vs. 90%) and ARR (two levels: 5% vs. 15%). The two between-subjects factors are mood (two levels: low mood vs. high mood) and numeracy (four

levels: poor vs. fair vs. good vs. excellent). If a higher RRR positively influences treatment perceptions then this would be indicated by an interaction between the B_RISK and the ARR: perceptions of treatment effectiveness should be boosted by the large RRR (i.e. 50%) at a small B_RISK (i.e. 30%) and a large ARR (i.e. 15%). If mood or numeracy moderated the impact of the RRR, then a three-way interaction would be expected between the B_RISK, the ARR and mood or numeracy, respectively. A number of covariates were included to control for confounding factors, including age, gender, education, and previous and past treatment.

2.7. ***Sample size calculation***

The power calculation was performed for determining the impact of the B_RISK on perceptions of effectiveness. A previous study on the B_RISK detected a small to medium effect size (partial η^2 (partial eta squared) = 0.0481) (Vogt et al., 2012). For factorial analysis of variance partial η^2 approximates the η^2 rules of thumb given as follows: small = 0.02, medium = 0.13, large = 0.26 (Cohen, 1988; Medical Research Council, 2013). The sample size for the current study was calculated conservatively with power at 0.90, with an alpha of 0.05 to detect a small to medium effect (partial η^2 = 0.0481) in perceived effectiveness of treatments using a main effects and interactions analysis of covariance (ANCOVA) with seven covariates. G*Power calculated a necessary sample size of 210 (Erdfelder, Faul, & Buchner, 1996).

2.8. ***Data preparation***

Prior to the analysis, the variables were examined for accuracy of data entry, outliers, violations of assumptions, and missing values. The minimum and maximum values, means and standard deviations of all the variables were found to be plausible apart from two participants. These participants had entered age as 1000 and 100, respectively, and responded to all questions with the minimum value; these participants were removed from the dataset. No significant univariate outliers were detected in the dataset (Tabachnick & Fidell, 2001).

2.8.1. Normality

Data was screened for normality examining skewness and kurtosis, both of which show zero in a perfectly normal distribution. The z score for skew and kurtosis was tested conservatively ($p < 0.001$), with scores < 3.29 considered normal (Tabachnick & Fidell, 2001).

None of the perceived effectiveness judgments of the vignettes showed skew: Vignette 5/10=50 ($z = 1.40$, $p > 0.001$), Vignette 15/30=50 ($z = -0.50$, $p > 0.001$), Vignette 5/30=17 ($z = 1.39$, $p > 0.001$), Vignette 15/90=17 ($z = 1.36$, $p > 0.001$), Vignette 5/90=6 ($z = 2.23$, $p > 0.001$), and Vignette 45/90=50 ($z = -2.87$, $p > 0.001$). Age ($z = -0.71$, $p > 0.001$), numeracy ($z = 1.57$, $p > 0.001$), and education ($z = -3.05$, $p > 0.001$) were not skewed either. None of the perceived effectiveness judgments of the vignettes showed kurtosis: Vignette 5/10=50 ($z = -1.89$, $p > 0.001$), Vignette 15/30=50 ($z = -1.86$, $p > 0.001$), Vignette 5/30=17 ($z = -1.84$, $p > 0.001$), Vignette 15/90=17 ($z = -1.79$, $p > 0.001$), Vignette 5/90=6 ($z = -1.63$, $p > 0.001$), and Vignette 45/90=50 (z = -

1.10, p > 0.001). Age (z = -1.09, p > 0.001), numeracy (z = -1.69, p > 0.001), and education (z = -1.10, p > 0.001) showed no kurtosis either. The remaining variables were dichotomous and normality does not apply.

2.8.2. Missing data

Answers for the perception about the effectiveness of talking-therapy were required and could not be skipped, resulting in no missing data. There were few missing data in the other variables. A cut-off that is often used to indicate needing to address missing data is 5% (Tabachnick & Fidell, 2001). There was one variable that exceeded this cut-off. This was the PHQ-4 score with 16 missing units (8% missing data). Because the variables enter the ANCOVA jointly, the combination of several variables with missing data, albeit each one with few missing data, can add up to cause a substantial loss of degrees of freedom and thus power. To prevent this, missing data points were replaced with the median score on age, gender, current treatment, past treatment, education, and mood. The median score on each variable was used to facilitate further analysis in the ANCOVA (e.g. keep variables dichotomous) (Tabachnick & Fidell, 2001).

2.8.3. Multicollinearity

One additional assumption for ANCOVA is the absence of multicollinearity. Multicollinearity is the presence of high correlations (i.e. > 0.9) between the CVs (Tabachnick & Fidell, 2001). If there is more than one CV and they are highly correlated they will cancel each other out of the equations. No multicollinearity was detected between gender, age, education, and present and past treatment.

3. Results

3.1. Overview

To gain an overview over the data, bi-variate correlations were computed for the six judgements of effectiveness and the covariates, demographic and psychosocial variables, as well as numeracy. See Table 3.1 for an overview.

3.1.1 Relation between six judgments of effectiveness

All judgements of effectiveness for the six different vignettes were significantly and positively correlated, with coefficients ranging from $r = 0.454$ to $r = 0.844$. This showed that participants' judgements of the vignettes was consistent to the extent that individuals who judged the effectiveness of talking-therapy as high in one vignette were likely to rate it as high in others as well.

3.1.2 Relation between covariates

Several significant correlations were detected among the covariates. Gender was correlated with numeracy ($r = -0.242$, $p = 0.001$), showing that male participants were more numerate. Education was correlated with numeracy ($r = 0.254$, $p < 0.001$), showing that numeracy increased with more education. Education was also correlated with mood ($r = -0.192$, $p = 0.008$), showing that those with more education had better mood. Currently receiving treatment for a CMHC was associated with past treatment for a CMHC ($r = 0.469$, $p = 0.001$).

3.1.3 Relation between six judgments of effectiveness and covariates

The correlation coefficients showed that there were few relationships between the covariates and the judgments of effectiveness. Indeed, out the 42

combinations, only four correlations were significant. Education was correlated with the effectiveness judgement of Vignette 5/30=17 (r = -0.159, p = 0.019), Vignette 5/90=6 (r = -0.168, p = 0.024), and Vignette 5/10=50 (r = -0.149, p = 0.039). In all these correlations, those with higher education judged the effectiveness of treatments as lower than those with lower education. Numeracy was also associated with the effectiveness judgement of 5/30=17 (r = -0.146, p = 0.041); more numerate participants judged talking-therapy in this vignette as less effective.

Table 3.1: Pearson's r correlations

	Age	Gender	Ctreat	Ptreat	Num	Edu	Mood	15/90=17	5/30=17	45/90=50	5/90=6	5/10=50	15/30=50
Age	1	-.055	.000	.025	-.050	.067	-.132	.076	.032	.036	.046	.023	.060
Gender	-.055	1	.092	.130	-.242**	.131	-.037	-.097	-.114	.005	-.129	-.123	-.069
Ctreat	.000	.092	1	.469**	-.043	-.088	-.061	-.027	-.055	.063	-.064	-.077	-.059
Ptreat	.025	.130	.469**	1	.077	.071	-.028	-.067	-.101	.053	-.121	-.087	-.075
Num	-.050	-.242**	-.043	.077	1	.254**	-.010	-.095	-.146*	.103	-.107	-.099	-.065
Edu	.067	.131	-.088	.071	.254**	1	-.192**	-.124	-.159*	.051	-.168*	-.149*	-.088
Mood	-.132	-.037	-.061	-.028	-.010	-.192**	1	-.063	-.046	-.124	-.036	-.074	-.091
15/90=17	.076	-.097	-.027	-.067	-.095	-.124	-.063	1	.673**	.723**	.844**	.603**	.674**
5/30=17	.032	-.114	-.055	-.101	-.146*	-.159*	-.046	.673**	1	.516**	.729**	.762**	.801**
45/90=50	.036	.005	.063	.053	.103	.051	-.124	.723**	.516**	1	.617**	.454**	.578**
5/90=6	.046	-.129	-.064	-.121	-.107	-.168*	-.036	.844**	.729**	.617**	1	.665**	.709**
5/10=50	.023	-.123	-.077	-.087	-.099	-.149*	-.074	.603**	.762**	.454**	.665**	1	.767**
15/30=50	.060	-.069	-.059	-.075	-.065	-.088	-.091	.674**	.801**	.578**	.709**	.767**	1

Notes: **. Correlation is significant at the 0.01 level (2-tailed); *. Correlation is significant at the 0.05 level (2-tailed); Gender (male/female); Ctreat = Current Treatment (no/yes); Ptreat = Past Treatment (no/yes); Num = Numeracy; Mood (normal/low); Edu = Education (low/high) 15/90=17 (perceived effectiveness); 5/30=17 (perceived effectiveness); 45/90=50 (perceived effectiveness); 5/90=6 (perceived effectiveness); 5/10=50 (perceived effectiveness); 15/30=50 (perceived effectiveness).

*3.2. Hypothesis 1 – interaction between RRR * ARR*

Hypothesis 1: Talking-therapies for conditions with higher B_RISKs are perceived as less effective.

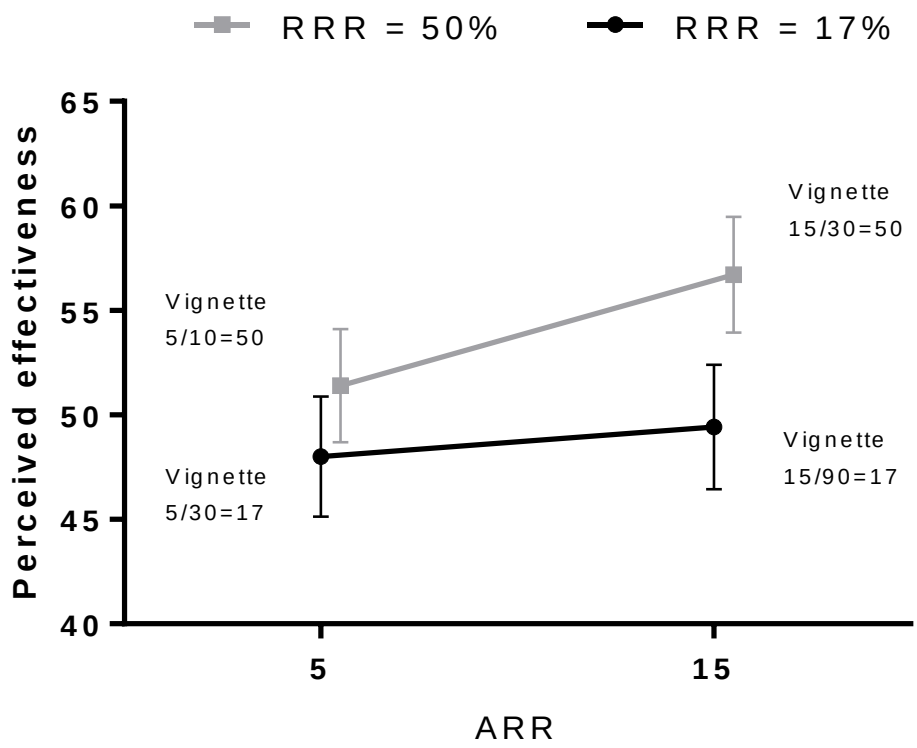
In order to test Hypothesis 1, a 2 x 2 x 2 x 4 mixed within-between subjects ANCOVA was run controlling for age, gender, education, previous and past treatment. The two-way interaction between the ARR and the RRR [$F(1, 191) = 5.292$, $p = 0.023$, partial $\eta^2 = 0.027$, Table 3.2] was significant. The ANCOVA also showed that both, the main effect for the RRR [$F(1, 191) = 29.666$, $p < 0.001$, partial $\eta^2 = 0.134$, Table 3.2] and for the ARR [$F(1, 191) = 5.363$, $p = 0.022$, partial $\eta^2 = 0.027$, Table 3.2] were significant. However, these main effects are interpreted considering the interaction between these factors (Tabachnick & Fidell, 2001).

Examining the means of these factors more closely (Figure 3.1), the interaction suggests that the RRR has less influence on the perceived effectiveness of talking-therapy at a low ARR (i.e. 5%). Similarly, the interaction suggests that that the ARR has less influence on the perceived effectiveness of talking-therapy at a low RRR (i.e. 17%). These results can be accounted by considering the B_RISK, to the extent that a larger B_RISK decreases the perceived treatment effectiveness of the talking-therapy.

At a low ARR (i.e. 5%) the B_RISK of Vignette 5/30=17 is just 20% higher than that of Vignette 5/10=50. However, at a high ARR (i.e. 15%) the B_RISK of Vignette 15/90=17 is 60% higher than that of Vignette 15/30=50, relatively depressing the perceived effectiveness of Vignette 15/90=17. Similarly, at a

high RRR (i.e. 50%) the B_RISK of Vignette 15/30=50 is just 20% higher than that of Vignette 5/10=50. However, at a low RRR (i.e. 17%) the B_RISK of Vignette 15/90=17 is 60% higher than that of Vignette 5/30=17, relatively depressing the perceived effectiveness of Vignette 15/90=17. The results therefore suggest that the B_RISK negatively influences perceptions of the effectiveness of talking-therapy for CMHCs.

Figure 3.1: Interaction between the ARR * RRR (mean, standard error)



3.2.1. No moderating effects

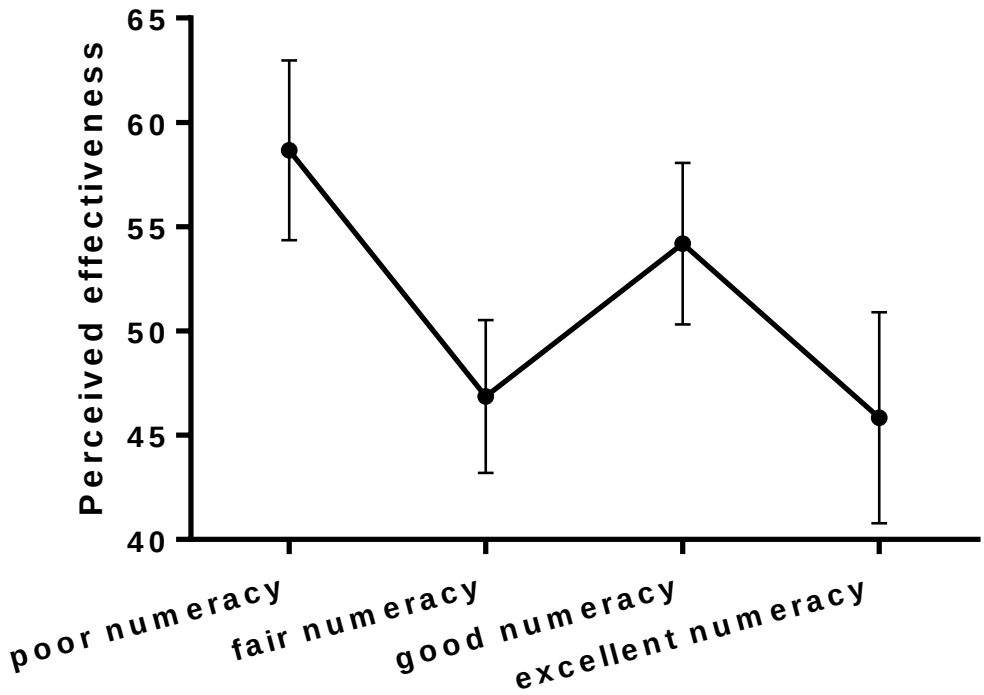
Neither the three-way interaction between the ARR, the RRR and numeracy [$F(3, 191) = 1.384$, $p = 0.249$, partial $\eta^2 = 0.021$, Table 3.2], nor the three-way interaction between the ARR, the RRR and mood [$F(1, 191) = 0.082$, $p = 0.775$, partial $\eta^2 < 0.001$, Table 3.2] was significant. There was also no other significant three-way interaction with a covariate and no significant two-way interaction (Table 3.2). This suggests that individuals were influenced by the

B_RISK and that this influence was not moderated by numeracy, mood, or any of the other covariates.

3.2.2. Numeracy influences perceived effectiveness

The main effect of numeracy was significant [F(3, 191) = 2.92, p = 0.035, partial η^2 = 0.044, Table 3.2]. Examining the means it appears that those with worse numeracy judge talking-therapies as more effective than those with better numeracy (Figure 3.2). Least significant difference (LSD) pairwise comparisons of the means show that indeed those with poor numeracy judged talking-therapies as more effective than those with fair numeracy [mean-difference = 11.81, p = 0.016] or excellent [mean-difference = -12.84, p = 0.047]. There were no further significant differences across different levels of numeracy.

Figure 3.2: Main effect of numeracy (mean, standard error)



3.2.3. Gender influences perceived effectiveness

Another main effect was significant. This was gender [$F(1, 191) = 4.12$, $p = 0.044$, partial $\eta^2 = 0.021$, Table 3.2]. The means show that women (mean = 47.91, SE = 2.60) judged talking-therapies as less effective than men (mean = 54.86, SE = 3.96).

Table 3.2: ANCOVA for RRR x ARR x Numeracy x Mood

Source	df	F	Sig.	partial η^2	Observed Power
RRR	1	29.666	<.001	.134	1.000
RRR * Num	3	.979	.404	.015	.264
RRR * Mood	1	.048	.828	<.001	.055
RRR * Ctreat	1	.507	.477	.003	.109
RRR * Ptreat	1	.049	.826	<.001	.056
RRR * Gender	1	1.851	.175	.010	.273
RRR * Age	1	1.718	.192	.009	.256
RRR * Edu	1	.635	.426	.003	.125
ARR	1	5.363	.022	.027	.635
ARR * Num	3	.745	.526	.012	.208
ARR * Mood	1	.370	.544	.002	.093
ARR * Ctreat	1	.452	.502	.002	.103
ARR * Ptreat	1	.121	.728	.001	.064
ARR * Gender	1	.119	.730	.001	.064
ARR * Age	1	.060	.807	<.001	.057
ARR * Edu	1	.013	.911	<.001	.051
RRR * ARR	1	5.292	.023	.027	.629
RRR * ARR * Num	3	1.384	.249	.021	.364
RRR * ARR * Mood	1	.082	.775	<.001	.059
RRR * ARR * Ctreat	1	<.001	.994	<.001	.050
RRR * ARR * Ptreat	1	.080	.777	<.001	.059
RRR * ARR * Gender	1	.238	.626	.001	.077
RRR * ARR * Age	1	.057	.811	<.001	.057
RRR * ARR * Edu	1	.064	.801	<.001	.057
Num	3	2.923	.035	.044	.689
Mood	1	.664	.416	.003	.128
Ctreat	1	.773	.381	.004	.141
Ptreat	1	.629	.429	.003	.124
Gender	1	4.122	.044	.021	.524
Age	1	.333	.564	.002	.089
Edu	1	3.415	.066	.018	.452
Error(RRR*ARR)	**191**				

Notes: Edu = Education; Ctreat = Current Treatment; Ptreat = Past Treatment; Num = Numeracy.

*3.3. Hypothesis 2 – Interaction between RRR * B_RISK*

Hypothesis 2: Talking therapies with higher ARRs are perceived as more effective.

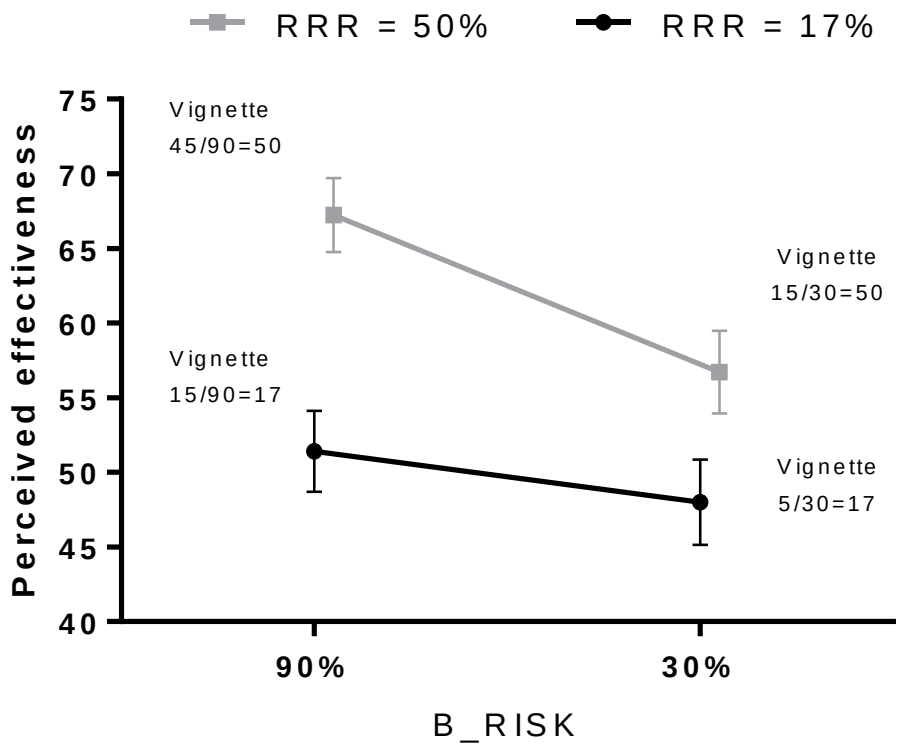
In order to test Hypothesis 2, a 2 x 2 x 2 x 4 mixed within-between subjects ANCOVA was run controlling for age, gender, education, previous and past treatment. The mixed ANCOVA showed that the two-way interaction between the B_RISK and the RRR [$F(1, 191) = 15.74$, $p < 0.001$, partial $\eta^2 = 0.076$, Table 3.3] was significant. The ANCOVA also showed that both the main effect of the RRR [$F(1, 191) = 165.462$, $p < 0.001$, partial $\eta^2 = 0.464$, Table 3.3] and that of the B_RISK [$F(1, 191) = 25.495$, $p < 0.001$, partial $\eta^2 = 0.118$, Table 3.3] was significant.

Examining the means of these factors more closely (Figure 3.3), the significant interaction suggests that the RRR has less influence on the perceived effectiveness of talking-therapy at a low B_RISK (i.e. 30%). Similarly, the significant interaction suggests that that the B_RISK has more influence on the perceived effectiveness of talking-therapy at a high RRR (i.e. 50%). These results can be accounted for by considering the ARR, to the extent that a larger ARR increases the perceived effectiveness of the talking-therapy.

At a low B_RISK (i.e. 30%) the ARR of Vignette 15/30=50 is just 10% higher than that of Vignette 5/30=17. However, at a high B_RISK (i.e. 90%) the ARR of Vignette 45/90=50 is 30% higher than that of Vignette 15/90=17, relatively boosting the perceived effectiveness of Vignette 45/90=50. Similarly, at a low RRR (i.e. 17%) the ARR of Vignette 15/90=17 is just 10% higher than that of Vignette 5/90=17. However, at a high RRR (i.e. 50%) the ARR of Vignette

45/90=50 is 30% higher than that of Vignette 15/30=50, relatively boosting the perceived effectiveness of Vignette 45/90=50. The results therefore suggest that the ARR positively influences perceptions of the effectiveness of talking-therapy for CMHCs.

Figure 3.3: Interaction between B_RISK * RRR (mean, standard error)



3.3.1. Moderating effects - numeracy

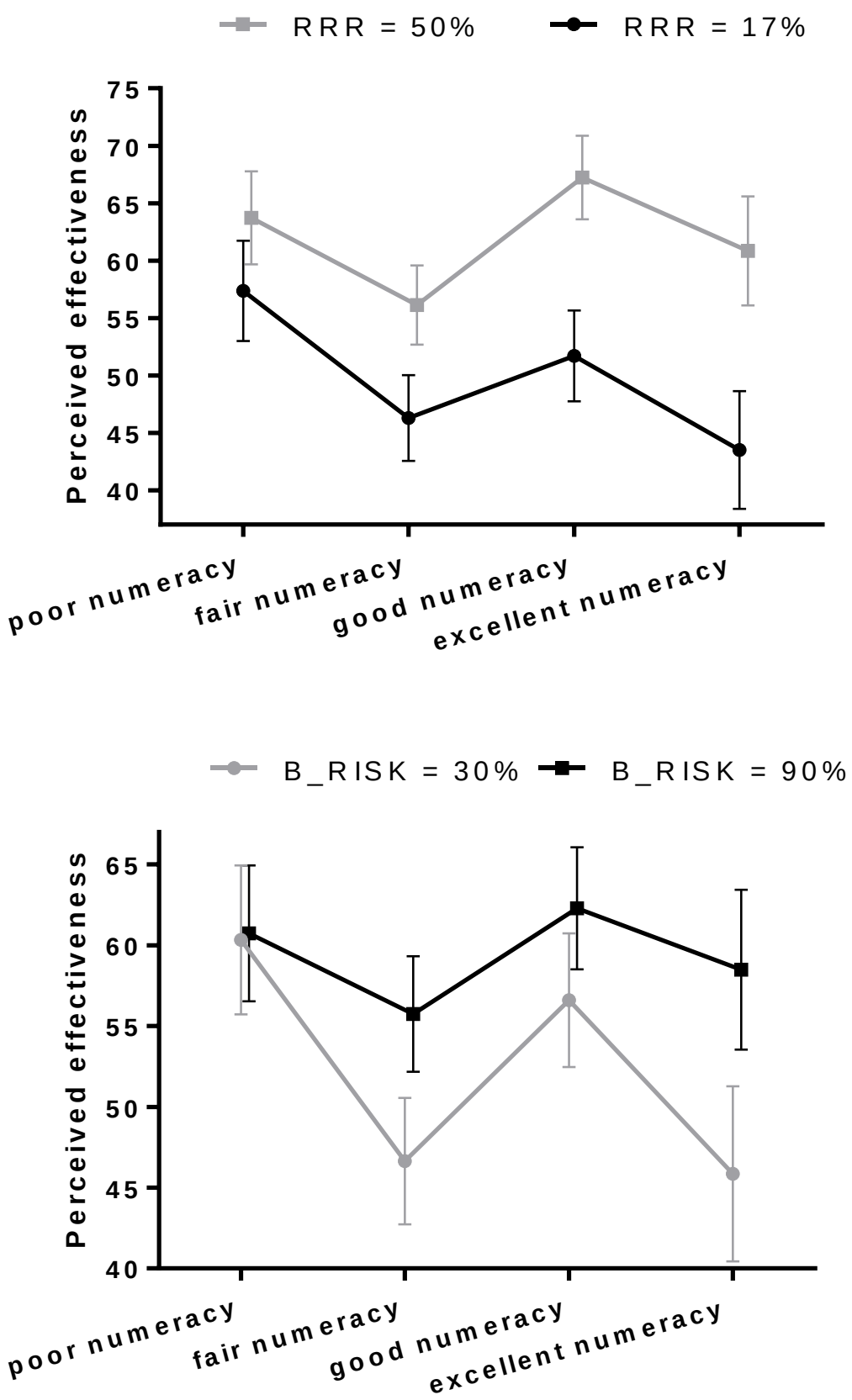
The three-way interaction between the B_RISK, the RRR, and numeracy was not significant [$F(3, 191) = 1.428$, $p = 0.236$, partial $\eta^2 = 0.022$, Table 3.3]. The three-way interaction between the B_RISK, the RRR, and mood was also not significant [$F(1, 191) = 0.066$, $p = 0.798$, partial $\eta^2 < 0.001$, Table 3.3]. There were no other three-way interactions between the B_RISK, the RRR and any of the covariates (Table 3.3). The results therefore suggest that individuals were influenced by the ARR, as indicated by the significant two-way interaction

between the ARR and the RRR, and that this influence is not moderated by numeracy, mood, or any of the other included covariates.

However, the two-way interactions between the B_RISK and numeracy [$F(3, 191) = 2.70$, $p = 0.047$, partial $\eta^2 = 0.041$, Table 3.3] and between the RRR and numeracy were significant [$F(3, 191) = 5.75$, $p = 0.001$, partial $\eta^2 = 0.083$, Table 3.3]. Because the two-way interaction between the B_RISK and the RRR was significant, indicating that the ARR influences results, neither the interaction between the B_RISK and numeracy, nor the interaction between the RRR and numeracy can be interpreted easily. That is, because the ARR is confounding these interactions.

The means of the interaction between the RRR and numeracy show that those with better numeracy differentiated more between talking-therapy with different RRRs, than those with poorer numeracy (Figure 3.4); with more numerate participants considering talking-therapy with a RRR of 17% less effective than a talking-therapy with a RRR of 50%. The means of the interaction between the B_RISK and numeracy show a similar pattern, in that those with better numeracy differentiated more between talking-therapy with different B_RISKs than those with poorer numeracy (Figure 3.4); with more numerate participants considering talking-therapy with a B_RISK of 30% as less effective.

Figure 3.4: Interactions between with Numeracy (mean, standard error)

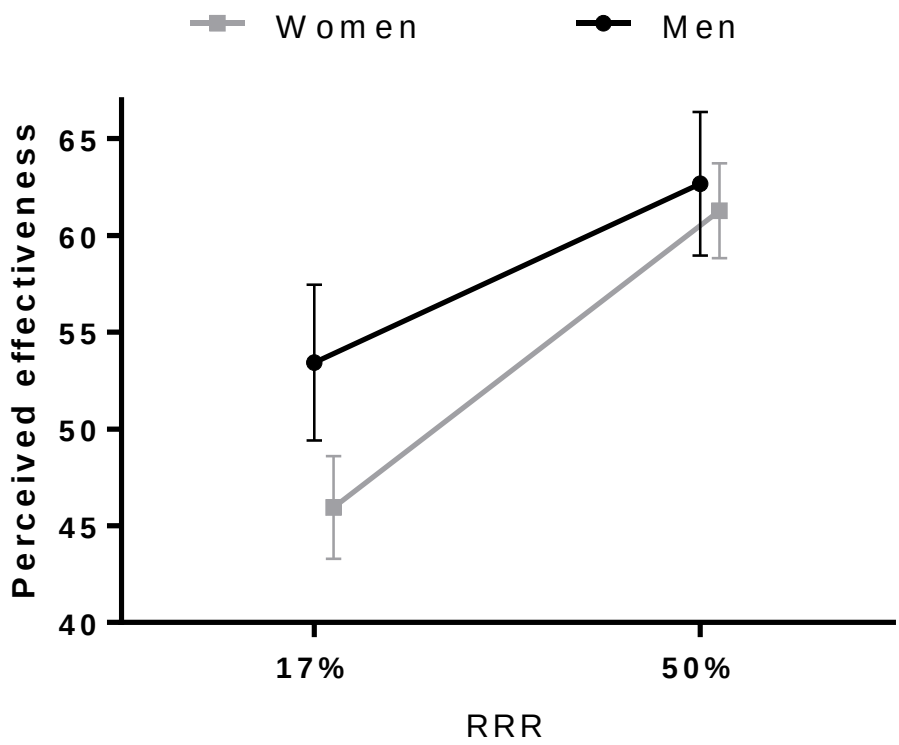


However, both the B_RISK and the RRR are confounded by the ARR. That is, talking-therapies with a higher B_RISK also have a higher ARR, and talking-therapies with a lower RRR also have a lower ARR. What these findings thus also show is that participants with worse numeracy respond less to differences in the ARR.

3.3.2. Moderating effects - gender

There was also a significant two-way interaction between the RRR and gender [$F(1, 191) = 8.816$, $p = 0.003$, partial $\eta^2 = 0.044$, Table 3.3]. The means show that women considered talking-therapy with a RRR of 17% as less effective than men, but not talking-therapy with a RRR of 50% (Figure 3.5).

Figure 3.5: Interaction between RRR * Gender (mean, standard error)



They may therefore be more sensitive to decreases in the RRR. However, the RRR in this analysis is confounded by the ARR, and thus the results also show that women consider treatments with a lower ARR as less effective than men; that is, they are more sensitive to decreases in the ARR than men. The interaction between the B_RISK and gender was not significant [$F(1, 191) = 1.818$, $p = 0.179$, partial $\eta^2 = 0.009$, Table 3.3].

Table 3.3: ANCOVA for RRR x B_RISK x Numeracy x Mood

Source	df	F	Sig.	partial η^2	Observed Power
RRR	1	165.462	<.001	.464	1.000
RRR * Num	3	5.746	.001	.083	.947
RRR * Mood	1	1.489	.224	.008	.229
RRR * Ctreat	1	1.687	.196	.009	.253
RRR * Ptreat	1	1.568	.212	.008	.238
RRR * Gender	1	8.816	.003	.044	.840
RRR * Age	1	.020	.888	<.001	.052
RRR * Edu	1	3.482	.064	.018	.459
B_RISK	1	25.495	<.001	.118	.999
B_RISK * Num	3	2.700	.047	.041	.650
B_RISK * Mood	1	.295	.588	.002	.084
B_RISK * Ctreat	1	1.958	.163	.010	.286
B_RISK * Ptreat	1	.592	.443	.003	.119
B_RISK * Gender	1	1.818	.179	.009	.269
B_RISK * Age	1	.013	.911	<.001	.051
B_RISK * Edu	1	1.323	.251	.007	.208
RRR * B_RISK	1	15.737	<.001	.076	.977
RRR * B_RISK * Num	3	1.428	.236	.022	.375
RRR * B_RISK * Mood	1	.066	.798	<.001	.058
RRR * B_RISK * Ctreat	1	1.867	.173	.010	.275
RRR * B_RISK * Ptreat	1	.391	.532	.002	.095
RRR * B_RISK * Gender	1	1.057	.305	.006	.176
RRR * B_RISK * Age	1	1.473	.226	.008	.227
RRR * B_RISK * Edu	1	1.316	.253	.007	.207
Num	3	2.207	.089	.034	.554
Mood	1	1.012	.316	.005	.170
Ctreat	1	.111	.739	.001	.063
Ptreat	1	.363	.547	.002	.092
Gender	1	2.032	.156	.011	.294
Age	1	.542	.463	.003	.113
Edu	1	1.935	.166	.010	.283
Error(RRR*B_RISK)	**191**				

Notes: Edu = Education; Ctreat = Current Treatment; Ptreat = Past Treatment; Num = Numeracy.

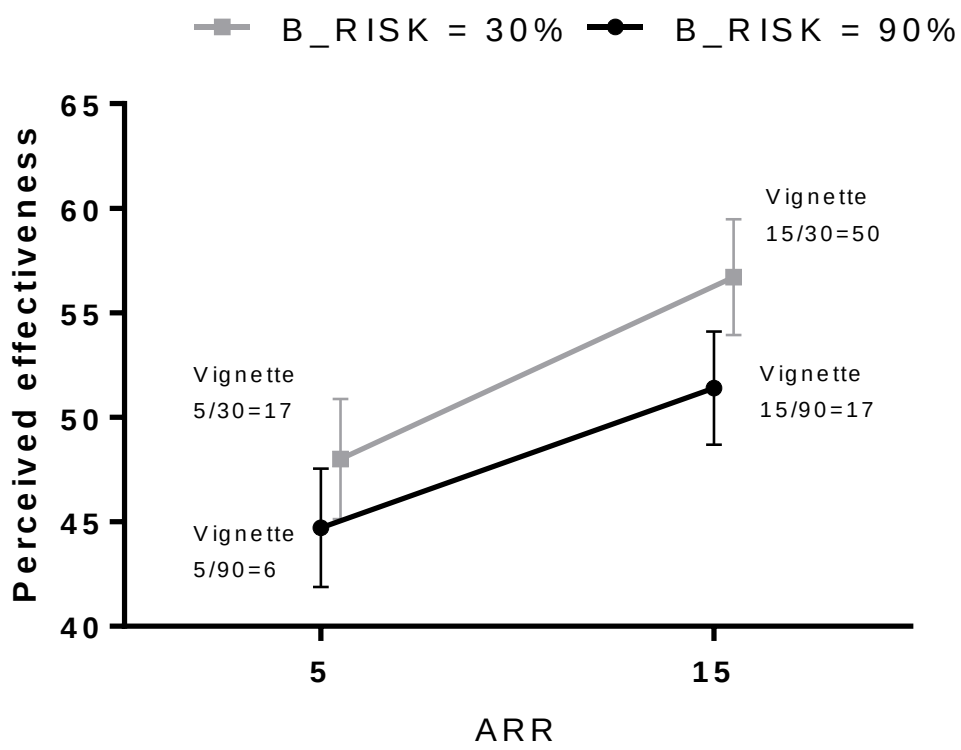
*3.4. Hypothesis 3 – No interaction between ARR * B_RISK*

Hypothesis 3: Talking therapies with higher RRRs are perceived as more effective.

In order to test Hypothesis 3, a 2 x 2 x 2 x 4 mixed within-between subjects ANCOVA was run controlling for age, gender, education, previous and past treatment. The mixed ANCOVA showed that the two-way interaction between the ARR and the B_RISK [$F(1, 191) = 1.858$, $p = 0.174$, partial $\eta^2 = 0.010$] was not significant (Table 3.4). This suggests that participants are not responding to different levels of the RRR.

The ANCOVA showed that the main effect of the ARR was significant [$F(1, 191) = 75.838$, $p < 0.001$, partial $\eta^2 = 0.284$, Table 3.4]. The means show that talking-therapies with a higher ARR are perceived as more effective than talking-therapies with a lower ARR (Figure 3.6). The ANCOVA also showed that the main effect of the B_RISK was significant [$F(1, 191) = 8.629$, $p = 0.004$, partial $\eta^2 = 0.043$, Table 3.4]. The means show that talking-therapies for conditions with a higher B_RISK are perceived as less effective than talking-therapies for conditions with a lower B_RISK (Figure 3.6).

Figure 3.6: No interaction between ARR * B_RISK (mean, standard error)



3.4.1. No moderating effects

There was no three-way interaction between the ARR, the B_RISK, and numeracy [$F(3, 191) = 1.064$, $p = 0.365$, partial $\eta^2 = 0.016$, Table 3.4] and no three-way interaction between the ARR, the B_RISK, and mood [$F(1, 191) = 0.039$, $p = 0.844$, partial $\eta^2 < 0.001$, Table 3.4]. There was also no other significant three-way interaction with a covariate and no significant two-way interaction (Table 3.4).

3.4.2. Gender influences perceived effectiveness

However, the between-subjects main effect of gender was significant [$F(1, 191) = 4.013$, $p = 0.043$, partial $\eta^2 = 0.021$, Table 3.4]. The means show that women (mean = 46.818, SE = 2.60) judged talking-therapies as less effective than men (mean = 53.599, SE = 3.955).

Table 3.4: ANCOVA for ARR x B_RISK x Numeracy x Mood

Source	df	F	Sig.	partial η^2	Observed Power
ARR	1	75.838	<.001	.284	1.000
ARR * Num	3	1.679	.173	.026	.435
ARR * Mood	1	.777	.379	.004	.142
ARR * Ctreat	1	.493	.483	.003	.108
ARR * Ptreat	1	1.237	.268	.006	.198
ARR * Gender	1	2.672	.104	.014	.370
ARR * Age	1	1.162	.282	.006	.189
ARR * Edu	1	1.045	.308	.005	.174
B_RISK	1	8.629	.004	.043	.832
B_RISK * Num	3	1.513	.213	.023	.395
B_RISK * Mood	1	.028	.867	<.001	.053
B_RISK * Ctreat	1	.018	.893	<.001	.052
B_RISK * Ptreat	1	.103	.749	.001	.062
B_RISK * Gender	1	.096	.757	.001	.061
B_RISK * Age	1	.044	.834	<.001	.055
B_RISK * Edu	1	.039	.843	<.001	.054
ARR * B_RISK	1	1.858	.174	.010	.274
ARR * B_RISK * Num	3	1.064	.365	.016	.285
ARR * B_RISK * Mood	1	.039	.844	<.001	.054
ARR * B_RISK * Ctreat	1	.479	.490	.002	.106
ARR * B_RISK * Ptreat	1	.158	.692	.001	.068
ARR * B_RISK * Gender	1	.349	.555	.002	.090
ARR * B_RISK * Age	1	.093	.761	<.001	.061
ARR * B_RISK * Edu	1	.023	.881	<.001	.053
Num	3	2.580	.055	.039	.628
Mood	1	.445	.506	.002	.102
Ctreat	1	.684	.409	.004	.130
Ptreat	1	.997	.319	.005	.168
Gender	1	4.013	.047	.021	.513
Age	1	.530	.467	.003	.112
Edu	1	3.418	.066	.018	.452
Error(ARR*B_RISK)	**191**				

Notes: Edu = Education; Ctreat = Current Treatment; Ptreat = Past Treatment; Num = Numeracy.

4. Discussion

4.1 Summary

The results showed that Hypothesis 1 was confirmed, suggesting that participants considered the B_RISK of the condition when they judged the effectiveness of a talking-therapy. The results also showed that Hypothesis 2 was confirmed, demonstrating that participants also considered the ARR of a talking-therapy when they judged its effectiveness at treating a CMHC. Finally, the results showed that Hypothesis 3 was not confirmed, the current study did not show that the RRR influenced participants' judgments of the effectiveness of a talking-therapy because the confounding variables ARR and B_RISK could not be excluded as alternative explanations.

There was no evidence that mood moderated the impact of the B_RISK, the ARR, or the RRR. There was some evidence that participants who failed to answer any of the numeracy questions were not responding to the RRR or the B_RISK. However, these findings were confounded by the ARR and an alternative explanation is that participants with worse numeracy skills are not sensitive to the ARR.

Finally, there was some evidence that women and those with better numeracy perceived talking-therapy as less effective than men and those with worse numeracy.

4.2 Hypothesis 1: B_RISK

The results showed that participants appeared to respond to the B_RISK of the condition when judging the effectiveness of a talking therapy. The current results therefore mirror those obtained in the study on judgments about medications for cancer (Vogt et al., 2012). It is thus only the second study that has documented that the B_RISK has an effect on perceptions of treatment effectiveness, which is independent from that of measures of treatment effectiveness. Previous studies, albeit reporting on the impact of the B_RISK on the perceived treatment effectiveness did not separate it from the confounding factors RRR and ARR (e.g. Fetherstonhaugh et al., 1997; Friedrich et al., 1999; Gyrd-Hansen et al., 2002; V. K. Smith & Desvousges, 1987). It is the first study that documented this effect for treatments aimed at CMHCs. The extent of the impact of the B_RISK was such that even talking-therapy with an ARR of 15 (Vignette 15/90=17) was not perceived as more effective than a talking-therapy with an ARR of only 5 (Vignette 5/30=17), because of the high B_RISK in the former.

An interpretation that could be drawn as a consequence of this study is that people who judge talking-therapies on the basis of the B_RISK are committing an error of judgment. They erroneously discredit talking-therapies for bad outcomes that are due to a naturally low rate of recovery and credit talking-therapies for high rates of recovery that are not caused by the treatment. This error has been named the ‘B_RISK integration error’ because participants integrate the B_RISK into their assessment of the treatment effect (Vogt et al., 2012).

4.2.1 Why are people influenced by the B_RISK when judging treatment effectiveness?

If individuals are making an erroneous decision when they evaluate the effectiveness of the talking-therapy on the basis of the B_RISK of the condition, the question arises as to the reason for this decision making error. A number of explanations may account for this findings.

“Persuasiveness of the RRR”

Some researchers have tried to explain the impact of the B_RISK in terms of the greater persuasiveness of the RRR as compared to the ARR on treatment perception (Baron, 1997). However, this explanation equates the impact of the B_RISK with that of the RRR, although they are conceptually and mathematically different factors. Thus, this explanation does not match the current findings which showed that the B_RISK had an effect on perceptions of treatment effectiveness that could not be accounted for by differences in the RRR.

“Psychological numbing”

Previous research on the B_RISK explained this apparent phenomenon as a result of "psychological numbing" or the “drop in the bucket” effect (Fetherstonhaugh et al., 1997). A tendency for individuals to rate an intervention saving a fixed number of lives to be less worth investing in, the greater the total number of lives that are not helped. In subsequent experiments participants were prompted to engage in more thoughtful processing about similar judgements, but these manipulations had no impact on "psychological numbing" (Friedrich et al., 1999). The researchers

interpreted this as indicating that rather than being the result of careless thinking, "psychological numbing" appears to be the result of a kind of reasoning that is meaningful to individuals.

While the above studies did not consider alternative explanations for their findings, namely the RRR, the processes underlying the B_RISK in the current study are consistent with the "psychological numbing" explanation. That is, talking-therapy for a population with a CMHC that has a large B_RISK might seem like a “drop in the bucket”, because many people have an adverse outcome despite treatment. The current study also showed that the impact of the B_RISK was not moderated by numeracy, which is consistent with the previous findings on “psychological numbing”, showing that it was not moderated by the degree of processing (Friedrich et al., 1999).

“Affect heuristic”

Another explanation for the impact of the B_RISK in the current study is what has been described as the “affect heuristic”. Early studies of risk perception found that, whereas risk and benefit tend to be positively correlated in the real world (e.g. a powerful technology/intervention also tends to have many adverse impacts, e.g. nuclear power), they are often negatively correlated in people’s minds (Fischhoff, Slovic, Lichtenstein, Read, & Combs, 1978). The reason for this negative correlation is postulated to be the “affect heuristic” (Slovic, Peters, Finucane, & MacGregor, 2005).

While exploring the negative correlation between the perceived risk and the perceived benefit of an intervention/behaviour it was found that it depended on the positive or negative affect associated with that intervention/behaviour

(Alhakami & Slovic, 1994). It suggested that people based their judgments of an intervention/behaviour not only on what they thought about it but also on how they felt about it. If feelings toward an intervention/behaviour were favourable, it was more likely that risks were judged as low and the benefits as high. If feelings toward intervention/behaviour were unfavourable, the opposite was more likely, risks were judged as high and the benefits as low. The “affect heuristic” postulates that affect comes prior to, and directs, judgments of risk and benefit (Slovic et al., 2005).

Based on the premise that peoples’ decisions depend on their affect, it was examined whether providing information about either the risks or benefits of an intervention/behaviour could change the affective view of the intervention/behaviour, which in turn might guide the perception of either benefits or risks (Finucane, Alhakami, Slovic, & Johnson, 2000). It was found that information stating the benefits as high for a technology such as nuclear power, lead to more positive affect, which, in turn, decreased the perceived risk of nuclear power. Similarly, providing information to increase the perceived risk decreased the perceived benefit of a technology via more negative affect.

The current finding may thus be explained by changes in the affect, caused by presenting CMHCs with low and high B_RISKs. Presenting a talking-therapy for the CMHC with a high B_RISK might have made the affective view about using the talking-therapy less favourable, in turn, leading to a decrease in the perceived effectiveness of the talking-therapy. By contrast, presenting a talking-therapy for a CMHC with a low B_RISK might have made affective view about using the talking-therapy more favourable, in turn, leading to an increase in the perceived effectiveness of the talking-therapy. Unfortunately, affect

corresponding to each vignette was not measured in the current study and therefore it is not possible to explore this explanation further. Future studies should however consider measuring affect corresponding to the vignette.

"Outcome bias"

Another heuristic that may account for the underlying processes responsible for the impact of the B_RISK on perceptions of treatment effectiveness is the "outcome bias". The "outcome bias" is a phenomenon whereby individuals judge health-care decisions by their ultimate outcome and not by the information that informed the decision known at the time the decision was made (Baron and Hershey 1988). In one of their experiments, an imaginary surgeon had to decide whether or not to perform a risky operation and participants were informed about the probabilities of success for each option. Participants were then told about the outcome of the operation and asked to judge the quality of the surgeon's decision. The findings showed that participants were more likely to rate the quality of the surgeon's decision as poor when the outcome of the operation had been bad. In other words, participants incorporated information that was unknown at the time the imaginary surgeon made the decision. They were thus holding the surgeon responsible for events beyond his/her control.

The current findings may reflect a similar process whereby participants make their decision about the effectiveness of talking-therapy dependent on the overall likelihood of a good outcome. Other things being equal, this is invariably greater if the B_RISK is low. However, if this accounts for the processes underlying the impact of the B_RISK, it suggests a thinking error that is related to people's understanding of numeracy. In turn, this would suggest that the

impact of the B_RISK might be moderated by numeracy. However, this was not detected in the current study. Alternatively, the findings documented under the "outcome bias" may themselves be explained through the "affect heuristic", although no research was found to examine this possibility.

"Availability heuristic"

Rather than being based on the "affect heuristic", the "outcome bias" may also be a result of the "availability heuristic". The "availability heuristic" proposes that individuals base their decisions on information that is most available in memory (Tversky & Kahneman, 1974). For example, individuals tend to overestimate the chance of dying from plane accidents and underestimate the chance of dying from car accidents because the former are more vivid, more unusual, and more emotionally charged. The "availability heuristic" may also account for process underlying the "outcome bias". That is, participants' responses are most influenced by the information provided to them last (i.e. whether surgery resulted in a good or bad outcome).

The "availability heuristic" is consistent with the current findings if the information most available in memory for the current participants was the remaining risk in the treatment group. The risk remaining in the treatment group is determined by the B_RISK and the risk reduction by the treatment. Unfortunately, the current study was not designed to examine memory content to evaluate this possible explanation.

It is possible that the "availability heuristic" has even a stronger impact on the judgement of treatment effectiveness outside of this lab-based experiment, that is, in the real world. Most individuals will not know the effectiveness of talking-

therapy in terms of the ARR or the RRR, or any other measure. The most likely information available to individuals is the risk remaining following treatment, as they observed friends or family going through treatment. As approximately 50% of those who had talking-therapy will not be in remission following treatment, treatments for CMHCs may be judged as ineffective, because the large proportion of people coming out of treatment might be salient compared to other health conditions. Media in film and writing may contribute to this, often portraying mental illness as untreatable (Wahl, 1997).

4.3 Hypothesis 2: ARR

The study also showed that participants judged those talking-therapies as more effective that had a higher ARR. In other words, participants were sensitive to differences in the ARR and thus the risk reduction provided. There is a lot of research describing that people respond to changes to the ARR (Akl et al., 2011; Harmsen et al., 2012). Furthermore, a review has looked specifically at how the format of presenting the ARR influences participants' judgment of the effectiveness of the treatments (Covey, 2007). The review found that often studies present the ARR without the B_RISK. This way of presenting information about treatments is deemed to be less informative for participants, in that, for example, it makes it more difficult for the reader to comprehend what the information actually means (Gigerenzer et al., 2007; Stovring, Gyrd-Hansen, Kristiansen, Nexoe, & Nielsen, 2008). In the current study the ARR was presented in conjunction with the B_RISK, thus ensuring that people were best able to comprehend the information.

When the ARR is presented alongside the B_RISK, responses to the ARR are potentially confounded by the B_RISK and the RRR, which can be extrapolated from the ARR and the B_RISK. To date, the only study that had previously documented the impact of the B_RISK on judgements of effectiveness while controlling for the ARR and the RRR and was conducted on cancer treatments, could not document that the responses to the ARR were independent of the B_RISK (Vogt et al., 2012). By contrast, the current study was designed to be able to avoid the possibility that the B_RISK could provide an alternative explanation for the impact of the ARR on treatment perceptions. The findings of the current study show, for the first time, that the ARR influences treatment perceptions independent of the B_RISK and the RRR.

4.4. Hypothesis 3: RRR

The findings also demonstrated that people were not sensitive to changes in the RRR. This findings is somewhat counterintuitive in light of the extensive literature accrediting the RRR more persuasiveness than the ARR (Akl et al., 2011). How can this finding be explained? There are several possible explanations.

First, many of the studies demonstrating the relative persuasiveness of the RRR are designed in a way so that participants are explicitly provided with the RRR. For example “Among those who take the pills, there will be a 33% reduced risk of heart disease during the next 10 years” (Carling et al., 2008). This way of presenting information may include reference to the B_RISK, but the B_RISK is often not provided (Covey, 2011)(e.g. Carling et al., 2008). Thus, participants in studies that have demonstrated the persuasiveness of the RRR

are often only looking at the impact of the explicitly reported RRR. By contrast, the RRR was not reported in the current study. Instead, participants could derive the RRR by looking at the proportional reduction of risk using the icon arrays that were provided. Had the RRR had been explicitly reported in the current study the impact of the RRR treatment perceptions might have been detectable. It has been shown that explicitly presenting the RRR leads to participants being more influenced by the size of the RRR (Gyrd-Hansen, Kristiansen, Nexøe, & Nielsen, 2003). The ARR was explicitly reported and an impact of the ARR was detected (see Hypothesis 2). The reason for not explicitly reporting the RRR was that communication guidelines recommend against this because it can make interventions seem more effective than they really are (Gigerenzer et al., 2007).

Second, detecting a unique impact of the RRR relied on participants being influenced by a difference in the RRR of 22% (i.e. 22% = (50% - 17%) - (17% - 6%); see Table 2.6). If the difference had been larger, for example 50%, than participants' responses to the RRR may have been more likely detect.

Third, the RRR has little or no influence on perceptions of treatment effectiveness. The size of the effect of the RRR, as provided by the interaction between the ARR and the B_RISK, was partial $\eta^2 = 0.01$, yielding an observed power of 0.274. A significant difference would have been detected for this effect size at a power of 0.80 had the sample size been 779. This might suggest that the findings from studies detecting such an impact may have been confounded by the B_RISK. However, in many studies the RRR was detected in which the B_RISK was not presented.

In summary, the absence of evidence to support the hypothesis that the RRR influences participants' perceptions of the effectiveness is likely to be due to a combination of not explicitly presenting the RRR, a too small difference in the RRR across the vignettes, and a lack of power. Thus, the design of the study rather than the RRR not influencing people's perceptions of the effectiveness of talking-therapy is likely to account for the current results.

4.4 Numeracy

4.5.1 Hypothesis 1: B_RISK

Numeracy was not found to moderate the impact of the B_RISK, as defined by a non-significant three-way interaction between the ARR, the RRR, and numeracy. This is consistent with previous research (i.e. Study 1 and Study 2) not detecting such a moderating impact (Vogt et al., 2012). In other words, the B_RISK seems to influence people's perceptions of treatment effectiveness if they have worse or better numeracy. This suggests that this impact is independent of mathematical understanding, making it unlikely that presenting the information in a numerically more accessible way, if such a way could be found, would change the results. The findings are consistent with the literature reviewing commonly reported heuristics (e.g. "loss aversion", "availability heuristic"), showing that these heuristic are often independent from cognitive ability (Stanovich & West, 2008).

An alternative explanation is that the lacking three-way interaction is a result of lacking power to detect such a difference. The effect size of the non-significant three-way interaction between the ARR, the RRR, and numeracy was partial $\eta^2 = 0.02$, yielding an observed power of 0.36. A significant difference would have

been detected for this effect size at a power of 0.80 had the sample size been 539. However, the results suggest that a moderating impact of numeracy on the impact of the B_RISK would be small.

There was a main effect of numeracy when examining Hypothesis 1, suggesting that those who were least numerate were judging talking-therapies as more effective. The results are consistent with those obtained from previous research (Study 1, Vogt et al., 2012), which showed that numeracy had a main effect on perceptions of treatment effectiveness; showing that those with poorer numeracy judged treatments as more effective than those with better numeracy. Some research has also found that those with poorer numeracy are more likely to opt for treatment (Carling et al., 2008). However, the perceived effectiveness and opting for treatment are different concepts, with the later being dependent on a number of other factors, such as perceived side-effects. Other research did not find that those with less numeracy were more likely to judge medical treatments as more effectiveness (Lipkus, Peters, Kimmick, Liotcheva, & Marcom, 2010). Interestingly, those with good numeracy perceived the treatment as effective as those with poor numeracy. This finding may be due to a sampling error.

4.5.2 Hypothesis 2: ARR

There was no evidence that numeracy moderated the impact of the ARR on treatment perceptions, as indicated by the non-significant interaction between the RRR, the B_RISK, and numeracy (see Table 3.3). However, numeracy was found to moderate the main effect of the RRR and the main effect of the B_RISK. There are at least three explanations for these findings.

First, the means suggest that those with lower numeracy were less responsive to differences in the RRR and the B_RISK. However, this explanation neglects the fact that the main effects of the RRR and the B_RISK were confounded by the differences in the ARR. That is, the main effects detected for the RRR and the B_RISK are caused, at least in part, by differences in the ARR. Furthermore, the current study found no evidence that the independent impact of the B_RISK was moderated by numeracy (as discussed under 4.5.1), and also did not detect an independent effect of the RRR on treatment perceptions (as discussed under 4.4).

Second, those with poor numeracy are responding less to differences in the ARR. The fact that the three-way interaction between the B_RISK, the RRR, and numeracy was not significant may be due to insufficient power to detect a significant effect. The effect size of the three-way interaction between the B_RISK, the RRR, and numeracy was partial $\eta^2 = 0.022$, yielding an observed power of 0.375. A significant difference would have been detected for this effect size at a power of 0.80 had the sample size been 489. Three previous studies are consistent this explanation .One showed that individuals who were more numerate were more responsive to differences in the ARR, with no difference detected with regards to the B_RISK (see Study 1 Vogt et al., 2012). It has also been documented that perceptions of treatment effectiveness in those with less numeracy were less responsive to the ARR (Lipkus et al., 2010). Another study found that those with more schooling (i.e. >10 years) were responsive to the size of the ARR, while those with less schooling (i.e. <10 years) were not (Gyrd-Hansen et al., 2003). The study also found that, the impact of the RRR on treatment perceptions was independent of education

(Gyrd-Hansen et al., 2003). While no impact of the education was detected in the current study, it may be a proxy for numeracy. Indeed, education was associated with numeracy in the current study.

Third, those with poor numeracy are not sensitive to changes in the ARR and the RRR. The fact that no three-way interactions with numeracy were detected may be due to lacking power. However, no independent effect of the RRR was found on treatment perceptions (as discussed under 4.4), suggesting it is unlikely that the RRR had an important function in the current study.

In summary, the evidence from the literature suggests that the impact of the ARR on treatment perceptions might be dependent on numeracy, but not that of the RRR or the B_RISK. Future research might further address the question of whether the ARR is moderated by numeracy with a sample.

Why would individuals with lower numeracy pay less attention to the treatment effect? There are at least two explanations and possibly a combination of these. First, the information may be too complex to be meaningfully understood by individuals who have poor numeracy. Indeed, having lower numeracy is associated with reduced information processing skills relevant to medical decision making, individuals with low numeracy levels are less likely to recall risk information and comprehend risk information (Zikmund-Fisher, Smith, Ubel, & Fagerlin, 2007). When individuals are asked about the information about medical treatments presented using the ARR, the RRR, and the B_RISK those with lower numeracy are less accurate in their judgment of the effectiveness of the medical treatment (Schwartz et al., 1997). Thus, in the

current study those with lower numeracy did not detect the lower effectiveness, that those with higher numeracy detected.

Second, the finding that participants did not respond to the size of the treatment effect may suggest that those who did not correctly answer numeracy questions did not actually read the information provided in the questionnaire. This may particularly apply to those who did not answer any numeracy question correctly. The numeracy questions that were used did require some understanding of chance and percentages but did not require participants to know complex mathematical concepts, such as algebra. As such, it is surprising that nearly a quarter of participants did not get any answer correct if they tried to do so. However, participants who failed to answer any questions correctly were not otherwise notable as outliers or provided implausible information. The hypothesis that all those who scored 0 on the numeracy measure did not read the questionnaire at all is therefore unlikely, although it may be true for some.

It is likely therefore that a combination of not understanding the information in the vignettes and not exerting sufficient effort into completing the questionnaire may be responsible for the finding that those with lower numeracy skills were less sensitive to differences in the treatment effect.

4.5 Mood

Although it was expected that participants' mood would have an impact on the judgements of effectiveness or moderate the impact of the B_RISK on judgements of effectiveness, no such relationships were detected. One would have predicted such relationships on the basis of a number of factors. For

example, as described earlier, the “affect heuristic” would postulate that participants with lower mood would have at least perceived all treatment as less effective than individuals with better mood (Finucane et al., 2000).

It is possible that mood did not have the expected impact on perceptions in the current study because compared to the normal population, the sample was skewed towards people with lower mood, with 75.8% of participants reporting symptoms of anxiety and sadness that indicate a CMHC. Thus, a ceiling effect may have been encountered.

It has been suggested that not all negative affect is equal, specifically showing that anxiety and sadness have different impacts on decision making (Raghunathan & Pham, 1999). This was investigated with regards to making decisions about the risks and benefits of gambling with money and with regards to making decisions about employment opportunities. The findings showed that sadness biases preferences toward high-risk/high-reward options, whereas anxiety biases preferences toward low-risk/low-reward options (Raghunathan & Pham, 1999). In the current study, a combined measure of sadness (PHQ-2) and anxiety (GAD-2) was used, namely the PHQ-4. Repeating the analysis by breaking up the combined PHQ-4 into sadness and anxiety produced virtually identical findings (results not reported in the results section but shown in Appendix 6). Thus, in the current study, even when considered separately, sadness and anxiety did not influence treatment perceptions and did not moderate the impact of the B_RISK.

4.6 Gender

In two of the three ANCOVA's (Table 3.2 and Table 3.4) it was found that men found talking-therapies more effective than woman and in one ANCOVA (Table 3.3), woman were more sensitive to differences in the treatment effect, perceiving vignettes with lower effectiveness as less effective than men, while being no different in their judgment of vignettes with higher effectiveness.

This finding is broadly consistent with a study of Irish Government employees in which men were found to have a higher intention to participate in counselling (Hyland, McLaughlin, et al., 2012). Similarly, a Dutch study found that men had a higher intention to seek professional help for a CMHC than women, although there were no significant differences in the attitude towards seeking help (Westerhof et al., 2008). Unfortunately, intention were not measured in the current study, which limits the comparison.

There are also studies that show no differences in attitude or intention to seek help for mental health problems between men and women (e.g. Vogel et al., 2005). However, it is generally assumed that women have a more positive attitude towards using talking-therapy, for example (Mackenzie et al., 2006, 2004). There is also a large number of studies showing that women have a higher intention to seek help for psychological difficulties (Addis & Mahalik, 2003; Leong & Zachar, 1999; Masuda, Suzumura, Beauchamp, Howells, & Clay, 2005; Rickwood & Braithwaite, 1994; Yeh, 2002).

The findings of the current study are therefore somewhat inconsistent with the larger literature. It is possible that the inconsistent findings are caused by cultural differences. European men may have a more favourable attitude

towards mental health services than non European men. Of the two studies that detected a more positive intention to seek help for psychological difficulties in men, one was conducted in Ireland and the other in Holland. By contrast, those studies detecting a more negative attitude and a lower intention to seek help among men were conducted in the US, Australia, Taiwan, and Japan. It is possible that culturally, talking-therapy is more acceptable for men in Europe, which is thus reflected in the current findings. Alternatively, the sample in the current study may have been self-selecting, reflecting a more positive attitude towards talking-therapy as compared to the general population. After all, participants were recruited into the study by placing advertisements about talking-therapy. It is possible though that only men who had a very positive attitude responded to the advertisement. This may be reflected in the lower proportion of men that was recruited into the current study.

It is also possible that the more positive beliefs about the effectiveness of talking-therapies among men was related to presenting information using icon-arrays. Receiving and processing this information could have impacted on participants. By contrast, in the above mentioned studies participants were simply asked about talking-therapies without having to first process numerical information. It is possible that men appreciated such information more, translating into higher perceptions of effectiveness.

4.7 Recommended decision making

The recommended gold-standard for judging treatments is that the decision should be based exclusively on measures of risk difference (i.e. ARR, odds ratio, number needed to treat) (Baron, 1997; Fetherstonhaugh et al., 1997;

Gigerenzer et al., 2007). Among health professionals the NNT is sometimes preferred because it frames the number of individuals at risk saved as the number of people that need to be treated to see a benefit from the treatment in one person (Cook & Sackett, 1995). As the NNT is the mathematical inverse of the ARR (i.e. 1/ARR=NNT), a treatment that has an ARR of 50% then has a NNT of 2; meaning that two people need to be treated for one person to benefit from the treatment. The advantage of focussing on the actual number of lives saved is that it is mathematically independent of the size of the B_RISK. For an individual at risk, a NNT of 2 can be interpreted as reflecting that there is a one in two chance that he/she will benefit from the treatment. Alternatively, a NNT of 10 (i.e. ARR = 10%) can be interpreted as reflecting that there is a one in 10 chance that a person will benefit from the treatment. Even though the person has a one in 10 chance of benefiting from the treatment, the person's absolute chance of a good outcome following treatment may be much higher, for example 90%, if the B_RISK is low (i.e. 20%).

The odds-ratio is another measure of treatment effectiveness. However, it is hard to interpret and thus not normally used to communicate the effectiveness of treatments (Deeks, Higgins, & Altman, 2008). However, the odds-ratio is commonly used in the analysis of randomised controlled trials of treatments for dichotomous outcomes, which are used to establish whether a treatment is better than standard care (Deeks et al., 2008). The odds-ratio is based on the frequencies of a 2 x 2 contingency table and is an index of the relationship between two inherently dichotomous variables (e.g. cured vs. uncured). Suppose a clinical trial for the effectiveness of a talking-therapy for depression had two conditions: a group that received talking-therapy and a group that

received care as usual, with 100 subjects each. The study reported that four subjects in the treatment condition were cured, whereas only two subjects in the control group were cured. These data are presented in a 2 x 2 contingency table (Table 4.1).

Table 4.1: 2 x 2 odds-ratio for B_RISK = 98% and ARR = 2%

	Cured		
	Yes	No	**Total**
New talking-therapy	4	96	**100**
Care as usual	2	98	**100**
Total	**6**	**194**	**200**

Note: cells are labelled a, b, c, d, clockwise starting in the top-left.

The cells have been labelled a, b, c, d, reading from left to right, top to bottom. The odds-ratio is calculated as follows (Lipsey & Wilson, 2001): OR = ad/bc = 4*98/96*2 = 2.0417. To establish whether this odds-ratio represents a statistically significant difference, the confidence interval (CI) can be calculated. For the present example the 95% CIs is 0.3654 to 11.4081 (Wilson, 2001), suggesting that the difference is not significant. The baseline risk in this example is 98%, the ARR is 2%, and the RRR is also 2%.

To demonstrate that the odds-ratio is independent of the B_RISK, in the next example the B_RISK is decreased to 4% keeping the ARR at 2%; this changes the RRR to 50%. This changes the cells in the 2 x 2 table (Table 4.2) and the calculation of the odds as follows: OR= ad/bc = 98*4/2*96 = 2.0417.

Table 4.2: 2 x 2 odds-ratio for B_RISK = 4% and ARR = 2%

	Cured		
	Yes	No	**Total**
New talking-therapy	98	2	**100**
Care as usual	96	4	**100**
Total	**6**	**194**	**200**

Note: cells are labelled a, b, c, d, clockwise starting in the top-left.

Despite the big change in the B_RISK, the result in terms of the odds-ratio remains the same. This example shows that the way in which the effectiveness of RCTs is calculated, by using the odds-ratio, depends on the actual difference that the treatment makes as compared to the control group, and not on the B_RISK or the RRR.

Other measures of effectiveness, such as the RRR, are mathematically depended on the B_RISK (Gigerenzer et al., 2007; Schünemann et al., 2008). That is, they suggest different levels of effect size depending on the size of the B_RISK, particularly, suggesting high effect sizes for low B_RISKs, and vice-versa. For this reason the RRR is often regarded as misleading (Fagerlin et al., 2011; Gigerenzer et al., 2007; Hoffrage, Lindsey, Hertwig, & Gigerenzer, 2000) and should not be used in the absence of presenting the B_RISK (Schünemann et al., 2008).

However, when comparing and combining studies across different risk groups as part of a meta-analysis, the RRR might still be useful. There has been a discussion about whether the actual treatment effect depends on the B_RISK of a population. Specifically, it has been argued that the same treatment may be more beneficial for people with a higher B_RISK, (i.e. more severe clinical

presentations / high risk groups) in terms of the ARR and less beneficial at a lower B_RISK in terms of the ARR (Weisberg, Hayden, & Pontes, 2009). When meta-analyses then combine data from low and high risk group, the mean value may then not be representative of those at low risk or those at high risk groups. By comparison, the RRR can be more stable across risk groups (Smeeth, Haines, & Ebrahim, 1999). A recommended approach for reviews is to present a variety of NNTs across a range of B_RISKs (Smeeth et al., 1999), as done in a review of oral anticoagulants to prevent stroke (Aguilar & Hart, 2005). However, a recent analysis of a meta-analysis on the effect of antidepressants on suicidality in children, did not find that the treatment effect, in terms of the ARR, depended on the B_RISK (Proschan, Brittain, & Fay, 2010).

4.8 Evidence for treatments of mental health conditions

Throughout the thesis reference was made to the natural rate of recovery in comparison to the effectiveness of talking-therapies, drawing upon data from either the Cochrane Collaboration or NICE guidelines (e.g. Bisson & Andrew, 2007; NICE, 2011). However, the data may be a simplification of the true rate of natural recovery from mental health conditions, which may be higher than that described in this thesis. Studying the untreated course of mental health conditions relies on using information from a variety of sources, including longitudinal studies prior to the development of treatments, wait-list controlled trials, or observational studies. These designs make it difficult to control for confounding factors, such as severity, or patients who do not seek treatment, who often experience less economical downfall (Coryell et al., 1995). For

example, an observational study that followed individuals with a recurrent depressive episode who did not seek treatment, found that 85% were in remission one year later (Michael A. Posternak et al., 2006). The study also found that those who sought help had lower recovery rates, more similar to those in wait-list controlled trials (M. A. Posternak & Miller, 2001). A recent meta-analysis using observational studies and wait-list controlled trials detected a recovery rate of untreated depression at 53% after one year (Whiteford et al., 2013). There is also data from naturalistic samples comparing patients who used antidepressants versus those who did not, showing that who did not use antidepressants did better one year later (Goldberg, Privett, Ustun, Simon, & Linden, 1998) and some research which suggests that a subset of patients may experience paradoxical effects (e.g. worsening of depression) from antidepressants when these are used for long periods (Fava, 2003). In summary, although the benefit of talking-therapy is not questioned per-se, it seems that at least some individuals may not be worse off without treatment; although the problem remains that it is not known beforehand who would be better off without treatment.

Also, questions have been asked about how evidence regarding effectiveness is established. First, this includes relying almost exclusively on the RCT for evaluating whether a treatment works with some suggesting the use of complementary effectiveness studies that allow for an evaluation of treatments that more closely resembles clinical practice (Blatt & Zuroff, 2005; Leichsenring & Leibing, 2007; Leichsenring, 2004). Specific issues include the reliance of RCTs on treatment manuals, including few co-morbidities, and the methodological inappropriateness of RCTs for long-term psychoanalytic

therapy. Second, it includes the relative neglect of the impact on functioning by focussing on symptoms, despite evidence that the correlation between symptoms and functioning is weak and often bidirectional (McKnight & Kashdan, 2009).

4.9 Weaknesses

The study had several weaknesses. First, the study does not allow firm conclusions about the relative strength of the B_RISK, the ARR, or the RRR. By selecting vignettes so that out of four vignettes, two vignettes were the same on one factor and two on another, the third factor varied across the four vignettes. It was this unrestrained variation that then allowed drawing inferences on whether this third unrestraint factor influenced perceptions about the effectiveness. As inferences were made on the basis of the size of the variation in the unrestraint factor, it was implicit that a larger variation in the unrestraint factor would lead to larger observable effect sizes. However, the variation could not be made of equal size across the three comparisons for mathematical reasons. For example, keeping the relation between the ARR and the RRR constant, the B_RISK varied by 40% (see Table 2.4). By comparison, keeping the relation between the B_RISK and the RRR constant, the ARR varied by 20% (see Table 2.5). Finally, keeping the relation between the ARR and the B_RISK constant, the RRR varied by 22% (see Table 2.6).

Second, to evaluate each of the three hypotheses, responses to four vignettes were analysed. In all, participants were asked to judge six vignettes. Some vignettes were thus used interchangeably for different analyses. The design was based on the approach successfully used to distinguish between absolute

and personal risk estimates (Mason, Prevost, & Sutton, 2008). The benefit of this approach is that it keeps participant effort at a minimum while maximising the use of the collected information. The alternative would have been to ask participants to judge 12 vignettes. The downside of using vignettes interchangeably is that it increases the risk of a confounding factor inherent in one of the vignettes to influence the results. However, it would not have been feasible for participants to judge 12 vignettes, as the risk of response fatigue was highly likely. It has been demonstrated that questions asked later in a long survey are often prone to more measurement error or misclassification (Egleston, Miller, & Meropol, 2011). It is also likely that extending the questionnaire and requiring 12 vignettes to be judged would have reduced response rates (Deutskens, Ruyter, Wetzels, & Oosterveld, 2004).

Third, intention to use talking-therapy were not measured and only perceptions about the effectiveness were assessed for each vignette. The reason was that it would have doubled the number of responses that participants would have had to make with regards to each vignette. Not having measured the intention to use talking-therapy limits the conclusions that can be drawn with regards to how relevant the current findings are at predicting whether or not people will make use of talking-therapies. However, beliefs about the effectiveness have been shown to be predictive of both intention to use talking-therapy, as well as, using taking-therapy in a previous study (Stecker et al., 2010). Similarly, there is a large body of evidence showing that intention is a determinant of behaviour (Webb & Sheeran, 2006).

Fourth, using colourful information for the icon-arrays may have affected individuals who are colour blind and thus introduced some measurement error.

The icon-arrays were displayed using the colours, green for those helped by talking therapy, blue for those recovering naturally, and red for those who do not recover. Studies use a variety of ways to display icon-arrays. Some use only black and white icons (Galesic et al., 2009), others used multicolour icons (Zikmund-Fisher et al., 2010). No evidence was identified in the literature that has evaluated the extent to which coloured or black/white icon-arrays influence decision making. However, using coloured icon-arrays, which in piloting of the current study were found to be easier to understand, may have limited the understanding of those that were colour blind. In the UK, colour blindness among woman is less than 0.5% and less than 10% in men (Swanson & Cohen, 2003). Given the demographic imbalance in the sample this meant that is likely that fewer than 2% of participants had some form of impairment in seeing colours, which may have impaired their responses. Unfortunately, no question was included to ask participants whether they had impaired vision. Future studies should include such a question.

Fifth, the type of icon used may have limited accuracy. Previous research has shown that the type of icon used in icon-arrays does not impact on the processing of risk information (Stone, Yates, & Parker, 1997). However, recent information shows that using restroom icons (i.e. ) results in better understanding of the risk information as compared to simpler icons, such as blocks or ovals, (Zikmund-Fisher et al., 2013). Therefore, the study may have become more sensitive to the changes in the B_RISK, the ARR, and the RRR, had restroom icons been used.

Sixth, the sample was collected entirely via the internet and may therefore not be entirely representative of the UK population. Using internet samples restricts the participants to those who have internet access. The pros and cons of using internet samples have been discussed elsewhere and even though internet-based studies are restrictive, face-to-face approaches place different restrictions on samples, for example they are usually obtained from specific locations or groups (Wright, 2005). By contrast, the current study was able to recruit participants from the whole of the UK and also age groups across the lifespan. However, the sample included more women than men. In the analysis gender was controlled for as a confounding factor and it was not detected to moderate the independent effects of the B_RISK and the ARR. Therefore, even though gender did have an impact on judgements of effectiveness, gender did not affect the three hypotheses. The sample also included fewer individuals from non-white ethnic groups (~6%) than would be expected from the general population in the UK, where 12% are from non-white ethnic groups (Office for National Statistics, 2011). The proportion of non-white individuals currently accessing the IAPT services ranges between10% and 16%, thus closely resembling the ethnicities in the UK (Glover, Webb, & Evison, 2010). Because of the large imbalance 6% to 94%, ethnicity was not included as a confounding variable (Tabachnick & Fidell, 2001).

Seventh, the circumstances under which the study attempted to investigate the factors that influence people's decisions are not reflective of what individuals encounter in the "real world" as they consider to use a talking-therapy for a CMHC. Patients considering to undergo cancer or cardiovascular treatment will now more frequently be offered information provided about the effectiveness of

the respective treatments. This is done to allow patients to make an informed decision about the treatment they are choosing (Wishart et al., 2010). By contrast, information about the effectiveness of talking-therapy is not offered routinely in the NHS. Therefore, the information shown to participants in the current study is unlikely to be currently presented to patients deciding to use talking-therapy. Nevertheless, it is postulated here that the judgement of the effectiveness of talking-therapy will inadvertently include common sense estimates of the B_RISK. Similarly, the somewhat artificial circumstances created for this study might become a common reality should a case be made in the NHS that making an informed decision about having talking-therapy should also include being informed about how effective it is likely to be (BPS, 2008, 2009).

Eight, the study was not designed to allow a follow-up to be conducted with the participants. This could have provided information about the stability of the impact of the B_RISK but also allowed to link participants' perceptions to future use of talking-therapy. Unfortunately, this was beyond the scope of this project.

Nine, participants were asked to state whether they were receiving or had received treatment for a mental health problem. Unfortunately, we did not distinguish whether this treatment included talking-therapy, pharmacotherapy, or both. Future studies should consider this.

Ten, of the 365 people who consented to the study, only 210 completed the items relating to the dependent variable. Those participants who completed the questionnaire may be different to those who did not, potentially introducing a bias. It is possible that they may be more favourable towards talking-therapy.

To explore this further, data that was collected but not included in the thesis is useful. The first page of the questionnaire measured constructs of the Theory of Planned Behaviour applied to the use of talking-therapy (see Appendix 4). Of the 365 that consented to the study, 348 completed questions about attitude the construct on the Theory of Planned Behaviour. When comparing those who completed the measure relating to the dependent variable (n = 202, mean = 4.803, SD = 1.690), which were presented on the second page of the questionnaire, to those who did not (n = 145, mean = 4.583, SD = 1.814) on the attitude towards talking-therapy, no significant difference was detected ($t(346) = 1.168$, $p = 0.244$). Thus, those who completed the questions relating to the dependent variable were not more favourable towards talking-therapy than those who dropped out earlier. It is possible that participants differed on another variable. For example, those who did not complete the questionnaire may have had lower numeracy and were deterred by the complexity of the vignettes. Unfortunately, it is not possible to explore this because numeracy was measured as one of the last constructs in the questionnaire and thus only completed by 202 participants.

4.10 Strengths

This study also had several strengths. First, the current study is the first to show that individuals decide on the effectiveness of talking-therapies for CMHCs on the basis of the B_RISK of the CMHCs.

Second, it is the first study to show an independent effect of the ARR when presenting treatment information with the B_RISK, controlling for differences in the RRR.

Third, the sample was large and heterogeneous on many accounts. Importantly, 75.8% of participants had a PHQ4 score of ≥ six. Therefore, the sample represented a group of individuals in which the majority presented with symptoms of anxiety and sadness that may warrant a diagnosis of a CMHC and thus represented candidates for talking-therapy. The findings are therefore not based on a student sample or a group of healthy adults. Indeed, evaluations from IAPT show that the sample is similar in terms of the case levels of individuals accessing IAPT, showing PHQ-9 scores beyond a level indicating a mood disorder in 72.5% of users and showing GAD-7 scores beyond a level indicating an anxiety disorder in 77.4% of users (Glover et al., 2010).

Fourth, criticisms about common heuristics include that evidence shows contradictory findings or their disappearance when natural frequency information rather than probabilistic information is used (Gigerenzer, 1991; Koehler, 1996). However, the information presented in the current study was specifically designed to present information in natural frequencies, so as to make it more likely that people could understand the information.

4.11 Clinical implications

The clinical implications of detecting that perceptions of the effectiveness of talking-therapies are influenced by the B_RISK are manifold.

4.11.1 For treatment uptake

Perceptions about the effectiveness are important determinants for the initial uptake of therapy (Stecker et al., 2010). Currently, patients are not provided

with information about the effectiveness of talking-therapy. Despite of this, their perceptions about the effectiveness may be influenced by the risk in the treatment group as individuals form their opinion about talking-therapy through hear-say or media outlets. Psychological services may thus start to present information about their effectiveness on printed leaflets or websites, to inform prospective patients about the benefit of treatment. Other services, such as stop smoking services, have adopted this approach (NHS-Smokefree, 2014).

There are also risks of presenting information about the ARR and the B_RISK. It is possible that by informing patients and other individuals about the impact of treatment and the B_RISK on the overall recovery rates, that they respond discouraged about the impact of talking-therapies. While the current study shows that a higher B_RISK, negatively, and higher ARR, positively, influence perceptions about the effectiveness of talking-therapies, some participants may hold overoptimistic perceptions about talking-therapies. In other words, when individuals learn about the ARR of talking-therapies and the B_RISKs, some may perceive them as less effective than they had perceived them before receiving the information. However, there is evidence from the promotion of stop smoking services that presenting smokers with the ARR of stop smoking services and the B_RISK of continuing to smoke increased attendance rates (Matcham, McNally, & Vogt, 2013).

4.11.2 For recovery rates

Perceptions about the effectiveness are also important determinants for the success of treatment once patients have engaged with therapy (Schindler, Hiller, & Witthöft, 2013). Patients who attend talking-therapy differ in the extent

to which they believe talking-therapy will help them overcome depression (Schindler et al., 2013). Those with a negative perception are 1.4 times more likely to drop out compared to those with a more positive perception. Asking patients about their perceptions about whether therapy will help them is therefore a valuable part in the assessment. Given the current findings, the conversation exploring the patient's perception should attend to the possibility that a negative perception is influenced by the B_RISK of CMHCs. If patients' expectations of treatment can be raised, this is likely to translate into lower drop-out rates and in turn to better outcomes.

It is important to keep a perspective about the actual clinical impact that beliefs about the effectiveness will have on the decision to use talking-therapy. Research about decisions for preventative treatments for heart disease showed that while important, the treatment effectiveness seems to have only a moderate influence on people's decisions about whether or not to use preventive medication (Harmsen et al., 2012). Other factors, such as personal or familial experience of the disease might be more important. This has been described as the 'single most important reason' (Czerlinski, Gigerenzer, & Goldstein, 1999). If patients' primary concerns are indeed not the effectiveness of a treatment, understanding their underlying reasons for decision-making is important, and the communication between patient and health professional should reflect this. Practitioners ought thus not focus solely on communicating effectiveness to their patients, but also on patient characteristics, including their values about seeking help, previous experiences of the individual patient (Gigerenzer et al., 2007), and intermediate therapy goals (Greenberger & Padesky, 1995). However, the treatment effectiveness may be something that

is amenable to change if a negative perception is at least partly influenced the by the B_RISK.

4.11.3 For gaining consent

Under professional guidelines, psychologists have to ensure that patients consent to the treatment, which requires patients to be informed about the likely outcomes of engaging in therapy, including its benefits. This is documented through the British Psychological Society's Code of Ethics and Consent (BPS, 2009). In this, the informed consent standard reads: *"(i) Ensure that clients, particularly children and vulnerable adults, are given ample opportunity to understand the nature, purpose, and anticipated consequences of any professional services or research participation, so that they may give informed consent to the extent that their capabilities allow."* Similarly, the British Psychological Society's Generic Professional Practice Guidelines states under informed consent about psychological activities that *"The psychologist should provide information about ... the benefits of this activity, either directly to the client in the case of assessment or intervention, or indirectly in the case of systemic intervention, or to potential theoretical advances in the case of research"* (BPS, 2008) . Currently, the information provided to patients about the likely success of therapy relies on providing patients with the dichotomous information that an intervention is effective, or not, for a specific condition, following evidence-based guidelines (e.g. NICE, 2009, 2011b). It does not, usually, involve more detailed information about the effect size (e.g. odds-ratio, ARR or RRR) or the B_RISK. If the standards of consent were to be tightened, as seen in other health conditions (Wishart et al., 2010), the effect size would

have to be communicated. This would require clinicians to take care that patients correctly interpret the treatment effect, without being influenced by the B_RISK.

4.11.4 For clinicians

Being more aware of the impact of the B_RISK on perceptions of treatment effectiveness may also be useful for individuals in dealing with treatment failures. It is unrealistic to expect that offering someone treatment will result in clinical success, as more than 50% of patients, that is, 1 in 2 patients are unlikely to go into remission for any given CMHC given the current state of treatments.

4.11.5 For referrers

Knowledge of the impact of the B_RISK may also be useful when dealing with referrers to psychological services. Anecdotal evidence from team meetings suggests that there are big differences in the frequency of referrals from individual health professionals, even in the same roles and services. Evidence from previous research showed that even those with medical training are likely to be influenced by the B_RISK when they evaluate the effectiveness of medical treatments (Vogt et al., 2012). It may be that when services are discussed with potential refers, the benefit of the talking-therapies are discussed by raising awareness that the B_RISK impacts on the overall proportion of people that can be expected to be in remission in any given CMHC following referral. There is also evidence that communicating the effectiveness of treatments for conditions with a high B_RISK (i.e. stop smoking services) to GPs by comparing it to other medical treatments using the

NNT increased referrals to stop smoking services (Vogt, Hall, Hankins, & Marteau, 2009). Presumably this was because GPs had underestimated the treatment effect relative to other treatments, because of the high B_RISK inherent to stop smoking services.

4.11.6 For commissioners

Knowledge of the B_RISK may also help in dealing with commissioners for psychological services. Often the way in which the effectiveness of the talking-therapies is described in reports only discusses the rate of recovery (Chan & Adams, 2014; Gyani, Shafran, Layard, & Clark, 2011). For example, in a report aimed at detailing how to improve recovery rates for IAPT services in the UK, the recovery rates for a depressive episode (40.3%), GAD (52.2%), PTSD (45.2%), OCD (42.7%), phobia (48.1%) are presented (Gyani et al., 2011). However, nowhere in the report is a reference made to what might have been expected had no treatment been offered, even though the rate of recovery is based in part on the B_RISK. In principle, it is therefore not at all possible to determine whether the IAPT services were effective or not. Presenting information about the effectiveness in this way might mean that commissioners respond adversely to the high B_RISK. Public health decisions have been reported to be affected by this kind of reasoning, as described by the reduced willingness to pay for interventions that reflected a “drop-in-the-bucket” (Fetherstonhaugh et al., 1997).

4.12 Future research

The research identified several avenues for further research, some of which are highlighted here. First, the role of affect as the underlying process for the

impact of the B_RISK should be explored. Second, does presenting information about the treatment effect of talking-therapy, with and without the B_RISK, increase perceptions of treatment effectiveness? Third, to identify tools to communicate the treatment effect of talking-therapy that avoid a negative influence of the B_RISK, possibly by comparing the treatment effect to that of other common treatments. There are also a number of issues that future research should pay attention to, including the use of restroom icons, controlling for colour blindness, and measuring intention to use talking-therapy. Services may also want to consider including a measure of the perceived outcome of therapy as part of routinely collected information when patient commence treatment, to be better able to explore this potential barrier.

4.13 Conclusion

Participants consider the B_RISK of the condition when they judge the effectiveness of a talking-therapy. The results also showed that participants consider the ARR of a talking-therapy when they judge the effectiveness of a talking-therapy. Finally, the results showed that the RRR did not seem to influence participants' judgments of the effectiveness of a talking-therapy in the current study.

References

Aday, L. A., & Andersen, R. (1974). A Framework for the Study of Access to Medical Care. *Health Services Research*, *9*(3), 208–220.

Addis, M. E., & Mahalik, J. R. (2003). Men, masculinity, and the contexts of help seeking. *The American Psychologist*, *58*(1), 5–14.

Aguilar, M. I., & Hart, R. (2005). Oral anticoagulants for preventing stroke in patients with non-valvular atrial fibrillation and no previous history of stroke or transient ischemic attacks. *Cochrane Database of Systematic Reviews*, *3*. doi:10.1002/14651858.CD001927.pub2

Ajzen, I. (1991). The theory of planned behaviour. *Organizational Behavior and Human Decision Processes*, *50*, 179 – 211. doi:10.1016/0749-5978(91)90020-T

Akl, E. A., Oxman, A. D., Herrin, J., Vist, G. E., Terrenato, I., Sperati, F., … Schünemann, H. (2011). Using alternative statistical formats for presenting risks and risk reductions. *The Cochrane Database of Systematic Reviews*, (3). doi:10.1002/14651858.CD006776.pub2

Alhakami, A. S., & Slovic, P. (1994). A psychological study of the inverse relationship between perceived risk and perceived benefit. *Risk Analysis*, *14*(6), 1085–1096.

Alonso, J., Codony, M., Kovess, V., Angermeyer, M. C., Katz, S. J., Haro, J. M., … Vilagut, G. (2007). Population level of unmet need for mental healthcare in Europe. *The British Journal of Psychiatry*, *190*(4), 299–306.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR.* Washington: American Psychiatric Publishing, Inc.

Ancker, J. S., Senathirajah, Y., Kukafka, R., & Starren, J. B. (2006). Design Features of Graphs in Health Risk Communication: A Systematic Review. *Journal of the American Medical Informatics Association*, *13*(6), 608–618. doi:10.1197/jamia.M2115

Andrews, G., Issakidis, C., & Carter, G. (2001). Shortfall in mental health service utilisation. *The British Journal of Psychiatry*, *179*(5), 417–425.

Armitage, C. J., & Conner, M. (2001). Efficacy of the Theory of Planned Behaviour: a meta-analytic review. *British Journal of Social Psychology*, *40*(4), 471–499. doi:10.1348/014466601164939

Bandura, A. (1986). *Social foundations of thought and action: a social cognitive theory*. New Jersey: Prentice Hall.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, *44*(3), 211–233.

Baron, J. (1997). Confusion of Relative and Absolute Risk in Valuation. *Journal of Risk and Uncertainty*, *14*(3), 301–309. doi:10.1023/A:1007796310463

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: The Guilford Press.

Becker, M. H., Maiman, L. A., Kirscht, J. P., Haefner, D. P., & Drachman, R. H. (1977). The Health Belief Model and prediction of dietary compliance: a field experiment. *Journal of Health and Social Behavior*, 348–366.

Bisson, J., & Andrew, M. (2007). Psychological Treatment of Post-traumatic Stress Disorder (PTSD). *The Cochrane Database of Systematic Reviews*, (3). doi:10.1002/14651858.CD003388.pub3

BPS. (2008). Generic Professional Practice Guidelines, 2nd Edition. The British Psychological Society.

BPS. (2009). Code of Ethics and Conduct. The British Psychological Society.

Bradley, C., Gamsu, D. S., Moses, J. L., Knight, G., Boulton, A. J., Drury, J., & Ward, J. D. (1987). The use of diabetes-specific perceived control and health belief measures to predict treatment choice and efficacy in a feasibility study of continuous subcutaneous insulin infusion pumps. *Psychology and Health*, *1*(2), 133–146.

Brown, S. M., Culver, J. O., Osann, K. E., MacDonald, D. J., Sand, S., Thornton, A. A., … Weitzel, J. N. (2011). Health literacy, numeracy, and interpretation of graphical breast cancer risk estimates. *Patient Education and Counseling*, *83*(1), 92–98. doi:10.1016/j.pec.2010.04.027

Cancer Research UK. (2013). Statistics and outlook for breast cancer□: Cancer Research UK□: CancerHelp UK. Retrieved 18 October 2013, from http://www.cancerresearchuk.org/cancer-help/type/breast-cancer/treatment/statistics-and-outlook-for-breast-cancer#trials

Carling, C., Kristoffersen, D. T., Herrin, J., Treweek, S., Oxman, A. D., Schünemann, H., … Montori, V. (2008). How should the impact of different presentations of treatment effects on patient choice be evaluated? A pilot randomized trial. *PloS One*, *3*(11), e3693. doi:10.1371/journal.pone.0003693

Carpenter, C. J. (2010). A Meta-Analysis of the Effectiveness of Health Belief Model Variables in Predicting Behavior. *Health Communication*, *25*(8), 661–669. doi:10.1080/10410236.2010.521906

Chan, S. W. Y., & Adams, M. (2014). Service Use, Drop-Out Rate and Clinical Outcomes: A Comparison Between High and Low Intensity Treatments in an IAPT Service. *Behavioural and Cognitive Psychotherapy*, 1–13. doi:10.1017/S1352465813000544

Chiles, J. A., Lambert, M. J., & Hatch, A. L. (1999). The Impact of Psychological Interventions on Medical Cost Offset: A Meta-analytic Review. *Clinical Psychology: Science and Practice*, *6*(2), 204–220.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Vol. 2). Hillsdale, NJ: Lawrence Earlbaum Associates.

Connelly, C. E. (1984). Compliance with outpatient lithium therapy. *Perspectives in Psychiatric Care*, *22*(2), 44–50.

Conner, M., & Norman, P. (1996). The role of social cognition in health behaviours. In M. Conner & P. Norman (Eds.), *Predicting health behaviour: Research and practice with social cognition models* (pp. 1–22). Maidenhead, BRK, England: Open University Press.

Cook, R. J., & Sackett, D. L. (1995). The number needed to treat: a clinically useful measure of treatment effect. *British Medical Journal*, *310*(6977), 452–454.

Covey, J. (2007). A Meta-analysis of the Effects of Presenting Treatment Benefits in Different Formats. *Medical Decision Making*, *27*(5), 638–654. doi:10.1177/0272989x07306783

Covey, J. (2011). The Effects of Absolute Risks, Relative Risks, Frequencies, and Probabilities on Decision Quality. *Journal of Health Communication*, *16*(7), 788–801. doi:10.1080/10810730.2011.561916

Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd, & A. R. Group (Eds.), *Simple heuristics that make us smart* (pp. 97–118). New York: Oxford University Press.

De Vries, M., Holland, R. W., Corneille, O., Rondeel, E., & Witteman, C. L. M. (2012). Mood effects on dominated choices: Positive mood induces departures from logical rules. *Journal of Behavioral Decision Making*, *25*(1), 74–81. doi:10.1002/bdm.716

Deeks, J. J., Higgins, J., & Altman, D. G. (2008). Analysing Data and Undertaking Meta-Analyses. In *Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series* (pp. 243–296). Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/9780470712184.ch9/summary

Deutskens, E., Ruyter, K. de, Wetzels, M., & Oosterveld, P. (2004). Response Rate and Response Quality of Internet-Based Surveys: An Experimental Study. *Marketing Letters*, *15*(1), 21–36. doi:10.1023/B:MARK.0000021968.86465.00

Dubow, E. F., Lovko Jr., K. R., & Kausch, D. F. (1990). Demographic Differences in Adolescents' Health Concerns and Perceptions of Helping Agents. *Journal of Clinical Child Psychology*, *19*(1), 44–54. doi:10.1207/s15374424jccp1901_6

Egleston, B. L., Miller, S. M., & Meropol, N. J. (2011). The impact of misclassification due to survey response fatigue on estimation and identifiability of treatment effects. *Statistics in Medicine*, *30*(30), 3560–3572. doi:10.1002/sim.4377

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, *28*, 1–11.

Estrada, C., Barnes, V., Collins, C., & Byrd, J. C. (1999). Health literacy and numeracy. *JAMA*, *282*(6), 527–527.

Fagerlin, A., Zikmund-Fisher, B. J., & Ubel, P. A. (2011). Helping Patients Decide: Ten Steps to Better Risk Communication. *Journal of the National Cancer Institute*, *103*(19), 1436–1443. doi:10.1093/jnci/djr318

Ferlie, E. B., & Shortell, S. M. (2001). Improving the Quality of Health Care in the United Kingdom and the United States: A Framework for Change. *Milbank Quarterly*, *79*(2), 281–315. doi:10.1111/1468-0009.00206

Fetherstonhaugh, D., Slovic, P., Johnson, S., & Friedrich, J. (1997). Insensitivity to the Value of Human Life: A Study of Psychophysical Numbing. *Journal of Risk and Uncertainty*, *14*(3), 283–300. doi:10.1023/A:1007744326393

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making*, *13*(1), 1–17.

Fischhoff, B., Slovic, P., Lichtenstein, S., Read, S., & Combs, B. (1978). How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. *Policy Sciences*, *9*(2), 127–152.

Fishbein, M. (1963). An investigation of the relationship between beliefs about an object and the attitude toward that object. *Human Relations*, *16*(3), 233–239.

Fishbein, M., Triandis, H. C., Kanfer, F. H., Becker, M., Middlestadt, S. E., & Eichler, A. (2001). Factors influencing behavior and behavior change. In A. Baum & T. A. Revenson (Eds.), *Handbook of health psychology* (pp. 3–17).

Floyd, D. L., Prentice-Dunn, S., & Rogers, R. W. (2000). A meta-analysis of research on protection motivation theory. *Journal of Applied Social Psychology*, *30*(2), 407–429.

Friedrich, J., Barnes, P., Chapin, K., Dawson, I., Garst, V., & Kerr, D. (1999). Psychophysical Numbing: When Lives Are Valued Less as the Lives at Risk Increase. *Journal of Consumer Psychology*, *8*(3), 277–299. doi:DOI: 10.1207/s15327663jcp0803_05

Galesic, M., Garcia-Retamero, R., & Gigerenzer, G. (2009). Using icon arrays to communicate medical risks: overcoming low numeracy. *Health Psychology*, *28*(2), 210–6. doi:10.1037/a0014474

Gigerenzer, G. (1991). How to Make Cognitive Illusions Disappear: Beyond 'Heuristics and Biases'. *European Review of Social Psychology*, *2*, 83–115.

Gigerenzer, G. (2003). *Reckoning with risk: learning to live with uncertainty*. London: The Penguin Press.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping Doctors and Patients Make Sense of Health

Statistics. *Psychological Science in the Public Interest*, *8*(2), 53–96. doi:10.1111/j.1539-6053.2008.00033.x

Glanz, K., Rimer, B. K., & Lewis, F. M. (2002). Health behavior and education. *Theory, Research and Practice. San Francisco: Wiley & Sons*.

Glover, G., Webb, M., & Evison, F. (2010). Improving Access to Psychological Therapies: A review of the progress made by sites in the first roll-out year. *North East Public Health Observatory*. Retrieved from http://www.iapt.nhs.uk/silo/files/iapt-a-review-of-the-progress-made-by-sites-in-the-first-roll8208-out-year.pdf

Godin, G., & Kok, G. (1996). The theory of planned behavior: A review of its applications to health-related behaviors. *American Journal of Health Promotion*, *11*(2), 87–98.

Grant, S., Aitchison, T., Henderson, E., Christie, J., Zare, S., McMurray, J., & Dargie, H. (1999). A comparison of the reproducibility and the sensitivity to change of visual analogue scales, borg scales, and likert scales in normal subjects during submaximal exercise. *CHEST Journal*, *116*(5), 1208–1217. doi:10.1378/chest.116.5.1208

Greenberger, D., & Padesky, C. A. (1995). *Mind over mood: Change how you feel by changing the way you think*. New York: The Guilford Press.

Gulliver, A., Griffiths, K. M., & Christensen, H. (2010). Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. *BMC Psychiatry*, *10*(1), 113. doi:10.1186/1471-244X-10-113

Gulliver, A., Griffiths, K. M., Christensen, H., & Brewer, J. L. (2012). A systematic review of help-seeking interventions for depression, anxiety and general psychological distress. *BMC Psychiatry*, *12*(1), 81.

Gyani, A., Shafran, R., Layard, R., & Clark, D. (2011). *Enhancing Recovery Rates in IAPT Services: lessons from analysis of the Year One data*. Retrieved from http://eprints.lse.ac.uk/47486/

Gyrd-Hansen, D., Kristiansen, I. S., Nexøe, J., & Nielsen, J. B. (2002). Effects of Baseline Risk Information on Social and Individual Choices. *Medical Decision Making*, *22*(1), 71–75. doi:10.1177/0272989x0202200107

Gyrd-Hansen, D., Kristiansen, I. S., Nexøe, J., & Nielsen, J. B. (2003). How Do Individuals Apply Risk Information When Choosing Among Health Care Interventions? *Risk Analysis*, *23*(4), 697–704. doi:10.1111/1539-6924.00348

Hammond, D., McDonald, P. W., Fong, G. T., & Borland, R. (2004). Do smokers know how to quit? Knowledge and perceived effectiveness of cessation assistance as predictors of cessation behaviour. *Addiction*, *99*(8), 1042–1048. doi:10.1111/j.1360-0443.2004.00754.x

Harmsen, C. G., Støvring, H., Jarbøl, D. E., Nexøe, J., Gyrd-Hansen, D., Nielsen, J. B., … Kristiansen, I. S. (2012). Medication effectiveness may not be the major reason for accepting cardiovascular preventive medication: A population-based survey. *BMC Medical Informatics and Decision Making*, *12*(1), 89. doi:10.1186/1472-6947-12-89

Harrison, J. A., Mullen, P. D., & Green, L. W. (1992). A meta-analysis of studies of the health belief model with adults. *Health Education Research*, *7*(1), 107–116.

Henshaw, E. J., & Freedman-Doan, C. R. (2009). Conceptualizing Mental Health Care Utilization Using the Health Belief Model. *Clinical*

Psychology: Science and Practice, *16*(4), 420–439. doi:10.1111/j.1468-2850.2009.01181.x

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science*, *290*(5500), 2261–2262.

Horne, R., & Weinman, J. (1998). Predicting treatment adherence: An overview of theoretical models. In Myers, Lynn B. & Midence, Kenny (Ed)., *Adherence to treatment in medical conditions* (pp. 25–50). Amsterdam: Harwood Academic Publishers.

Hunot, V., Churchill, R., Teixeira, V., & Silva de Lima, M. (2007). Psychological therapies for generalised anxiety disorder. *Cochrane Database of Systematic Reviews*.

Hyland, P. E., Boduszek, D., Shevlin, M., & Adamson, G. (2012). Investigating the Latent Structure of the Theory of Planned Behavior and its Effectiveness in Explaining Intentions to Participate in Counseling among a Sample of Police Officers. *Journal of Police and Criminal Psychology*, *27*(2), 129–140. doi:10.1007/s11896-012-9103-9

Hyland, P. E., McLaughlin, C. G., Boduszek, D., & Prentice, G. R. (2012). Intentions to participate in counselling among front-line, at-risk Irish government employees: an application of the theory of planned behaviour. *British Journal of Guidance & Counselling*, *40*(3), 279–299. doi:10.1080/03069885.2012.681769

Iles, V., & Sutherland, K. (2001). Organisational Change: A Review for Health Care Managers, Professionals and Researchers. London School of Hygiene and Tropical Medicine.

Kelly, G. R., Mamon, J. A., & Scott, J. E. (1987). Utility of the health belief model in examining medication compliance among psychiatric outpatients. *Social Science & Medicine*, *25*(11), 1205–1211.

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences*, *19*(1), 1–53.

Kroenke, K., Spitzer, R. L., & Williams, J. B. . (2001). The PHQ-9. *Journal of General Internal Medicine*, *16*(9), 606–613.

Kroenke, K., Spitzer, R. L., Williams, J. B., Monahan, P. O., & Lowe, B. (2007). Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine*, *146*(5), 317–325.

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2003). The Patient Health Questionnaire-2: validity of a two-item depression screener. *Medical Care*, *41*(11), 1284–1292.

Kroenke, K., Spitzer, R. L., Williams, J. B. W., & Löwe, B. (2009). An ultra-brief screening scale for anxiety and depression: the PHQ-4. *Psychosomatics*, *50*(6), 613–621. doi:10.1176/appi.psy.50.6.613

Layard, R. (2005). *Mental health: Britain's biggest social problem?* Paper presented at the No.10 Strategy Unit Seminar on Mental Health on 20 th January 2005. Retrieved from http://eprints.lse.ac.uk/47428/

Layard, R. (2012). Mental illness loses out in the NHS. *British Politics and Policy at LSE*. Retrieved from http://eprints.lse.ac.uk/48405/

Layard, R. (2013). Mental health: the new frontier for labour economics. *IZA Journal of Labor Policy*, *2*(1), 2. doi:10.1186/2193-9004-2-2

Layard, R., Clark, D., Knapp, M., & Mayraz, G. (2007). Cost-benefit analysis of psychological therapy. *National Institute Economic Review*, *202*(1), 90–98.

Leong, F. T., & Zachar, P. (1999). Gender and opinions about mental illness as predictors of attitudes toward seeking professional psychological help. *British Journal of Guidance and Counselling*, *27*(1), 123–132.

Lewis, C. E., Fein, R., & Mechanic, D. (1976). *A right to health: the problem of access to primary medical care*. New York: Wiley.

Li, C., Friedman, B., Conwell, Y., & Fiscella, K. (2007). Validity of the Patient Health Questionnaire 2 (PHQ-2) in Identifying Major Depression in Older People. *Journal of the American Geriatrics Society*, *55*(4), 596–602. doi:10.1111/j.1532-5415.2007.01103.x

Lipkus, I. M., Peters, E., Kimmick, G., Liotcheva, V., & Marcom, P. (2010). Breast Cancer Patients' Treatment Expectations after Exposure to the Decision Aid Program Adjuvant Online: The Influence of Numeracy. *Medical Decision Making*, *30*(4), 464–473. doi:10.1177/0272989x09360371

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General Performance on a Numeracy Scale among Highly Educated Samples. *Medical Decision Making*, *21*(1), 37–44. doi:10.1177/0272989X0102100105

Lipsey, M. W., & Wilson, D. B. (2001). *Practical Meta Analysis. Applied Social Research Methods*. Thousand Oaks, CA: Sage.

Löwe, B., Kroenke, K., & Gräfe, K. (2005). Detecting and monitoring depression with a two-item questionnaire (PHQ-2). *Journal of*

Psychosomatic Research, *58*(2), 163–171. doi:10.1016/j.jpsychores.2004.09.006

Löwe, B., Wahl, I., Rose, M., Spitzer, C., Glaesmer, H., Wingenfeld, K., … Brähler, E. (2010). A 4-item measure of depression and anxiety: validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. *Journal of Affective Disorders*, *122*(1), 86–95.

Mackenzie, C. S., Gekoski, W. L., & Knox, V. J. (2006). Age, gender, and the underutilization of mental health services: The influence of help-seeking attitudes. *Aging & Mental Health*, *10*(6), 574–582. doi:10.1080/13607860600641200

Mackenzie, C. S., Knox, V. J., Gekoski, W. L., & Macaulay, H. L. (2004). An Adaptation and Extension of the Attitudes Toward Seeking Professional Psychological Help Scale1. *Journal of Applied Social Psychology*, *34*(11), 2410–2433. doi:10.1111/j.1559-1816.2004.tb01984.x

Mason, D., Prevost, A. T., & Sutton, S. (2008). Perceptions of absolute versus relative differences between personal and comparison health risk. *Health Psychology*, *27*(1), 87–92.

Masuda, A., Suzumura, K., Beauchamp, K. L., Howells, G. N., & Clay, C. (2005). United States and Japanese college students' attitudes toward seeking professional psychological help. *International Journal of Psychology*, *40*(5), 303–313.

Matcham, F., McNally, L., & Vogt, F. (2013). A pilot randomized controlled trial to increase smoking cessation by maintaining National Health Service

Stop Smoking Service attendance. *British Journal of Health Psychology*. doi:10.1111/bjhp.12078/full

Mathers, C., Fat, D. M., & Boerma, J. T. (2008). *The global burden of disease: 2004 update*. New York: World Health Organization.

MayoClinic. (2013). Cancer survival rate: A tool to understand your prognosis - MayoClinic.com. Retrieved 18 October 2013, from http://www.mayoclinic.com/health/cancer/CA00049

McCormack, H. M., Horne, D. J., & Sheather, S. (1988). Clinical applications of visual analogue scales: a critical review. *Psychological Medicine*, *18*(4), 1007–1019.

McCormack, J. (2013). Cardiovascular Risk/Benefit Calculator. Retrieved 18 October 2013, from http://bestsciencemedicine.com/chd/calc2.html

McManus, S. (2009). *Adult psychiatric morbidity in England, 2007: results of a household survey*. London: National Centre for Social Research.

Mechanic, D. (1962). The concept of illness behavior. *Journal of Chronic Diseases*, *15*, 189.

Medical Research Council. (2013). FAQ/effectSize - MRC Cognition and Brain Sciences Unit - statistics Wiki. Retrieved 4 May 2014, from URL:http://imaging.mrc-cbu.cam.ac.uk/statswiki/FAQ/effectSize. Accessed: 2014-05-04. (Archived by WebCite® at http://www.webcitation.org/6PJy0gOcq)

Meltzer, P. B. (2000). The reluctance to seek treatment for neurotic disorders. *Journal of Mental Health*, *9*(3), 319–327.

Mental Health Policy Group. (2006). *The Depression Report: A New Deal for Depression and Anxiety Disorders*. London: The Centre for Economic Performance.

Mental Health Policy Group. (2012). *How mental illness loses out in the NHS*. London: The Centre for Economic Performance.

Michaelson, J. (2011). *Breast Cancer Outcome Calculator*. US: Laboratory for Quantitative Medicine, Massachusetts General Hospital. Retrieved from http://www.lifemath.net/cancer/breastcancer/outcome/index.php

Mills, J. L. (2010). *Does the Theory of Planned Behavior Predict Intentions to Seek Help for Suicidality?* (Thesis). Marshall University. Retrieved from http://mds.marshall.edu/etd/105/

Milne, S., Sheeran, P., & Orbell, S. (2000). Prediction and Intervention in Health Related Behaviour: A meta-analytic review of Protection Motivation Theory. *Journal of Applied Social Psychology*, *30*(1), 106–143.

NHS Choices. (2014). Retrieved 22 April 2014, from URL:http://www.nhs.uk/NHSEngland/AboutNHSservices/mentalhealthservices/Pages/Overview.aspx. Accessed: 2014-04-24. (Archived by WebCite® at http://www.webcitation.org/6P4gDsiNW)

NHS-Smokefree. (2014). Local NHS Stop Smoking Services. Retrieved 22 April 2014, from http://www.nhs.uk/smokefree/help-and-advice/local-support-services-helplines (Archived by WebCite® at http://www.webcitation.org/6P1le3vGI)

NICE. (2009). *Depression: the treatment and management of depression in adults* (CG90). London: National Institute for Health and Clinical Excellence.

NICE. (2011a). *Common mental health disorders: identification and pathways to care* (CG123). London: National Institute for Health and Clinical Excellence.

NICE. (2011b). *Generalised anxiety disorder and panic disorder (with or without agoraphobia) in adults: management in primary, secondary and community care.* (CG113). London: National Institute for Health and Clinical Excellence.

Office for National Statistics. (2011). 2011 Census: Aggregate data (England and Wales). UK Data Service Census Support. Retrieved from http://infuse.mimas.ac.uk

Oliver, M. I., Pearson, N., Coe, N., & Gunnell, D. (2005). Help-seeking behaviour in men and women with common mental health problems: cross-sectional study. *The British Journal of Psychiatry*, *186*(4), 297–301. doi:10.1192/bjp.186.4.297

Painter, J. E., Borba, C. P. C., Hynes, M., Mays, D., & Glanz, K. (2008). The Use of Theory in Health Behavior Research from 2000 to 2005: A Systematic Review. *Annals of Behavioral Medicine*, *35*(3), 358–362. doi:10.1007/s12160-008-9042-y

Pan, P. C., & Tantam, D. (1989). Clinical characteristics, health beliefs and compliance with maintenance treatment: a comparison between regular and irregular attenders at a depot clinic. *Acta Psychiatrica Scandinavica*, *79*(6), 564–570.

Pescosolido, B. A. (1992). Beyond rational choice: The social dynamics of how people seek help. *American Journal of Sociology*, 1096–1138.

Pescosolido, B. A., Gardner, C. B., & Lubell, K. M. (1998). How people get into mental health services: stories of choice, coercion and 'muddling through' from 'first-timers'. *Social Science & Medicine*, *46*(2), 275–286.

Proschan, M. A., Brittain, E. H., & Fay, M. P. (2010). Does treatment effect depend on control event rate? Revisiting a meta-analysis of suicidality and antidepressant use in children. *Clinical Trials*, *7*(2), 109–117.

Raghunathan, R., & Pham, M. T. (1999). All negative moods are not equal: Motivational influences of anxiety and sadness on decision making. *Organizational Behavior and Human Decision Processes*, *79*(1), 56–77.

Reyna, V. F., & Brainerd, C. J. (2007). The importance of mathematics in health and human judgment: Numeracy, risk communication, and medical decision making. *Learning and Individual Differences*, *17*(2), 147–159. doi:10.1016/j.lindif.2007.03.010

Rickwood, D., & Braithwaite, V. A. (1994). Social-psychological factors affecting help-seeking for emotional problems. *Social Science & Medicine*, *39*(4), 563–572.

Rickwood, D., Deane, F. P., Wilson, C. J., & Ciarrochi, J. (2005). Young people's help-seeking for mental health problems. *Advances in Mental Health*, *4*(3), 218–251.

Rickwood, D., & Thomas, K. (2012). Conceptual measurement framework for help-seeking for mental health problems. *Psychology Research and Behavior Management*, *5*, 173–183. doi:10.2147/PRBM.S38707

Rosenstock, I. M., & Kirscht, J. (1979). Why people seek health care. In G. Stone & N. Cohen (Eds.), *Health psychology*. San Francisco: Jossey-Bass Publishers.

Rosenstock, I. M., Strecher, V. J., & Becker, M. H. (1994). The Health Belief Model and HIV Risk Behavior Change. In R. J. DiClemente & J. L. Peterson (Eds.), *Preventing AIDS* (pp. 5–24). Springer US. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4899-1193-3_2

Sackett, D. L., Richardson, W., Rosenberg, W., & Haynes, R. (1997). *Evidence-based medicine: How to practice and teach EBM*. New York: Churchill Livingston.

Schindler, A., Hiller, W., & Witthöft, M. (2013). What Predicts Outcome, Response, and Drop-out in CBT of Depressive Adults? A Naturalistic Study. *Behavioural and Cognitive Psychotherapy*, *41*(03), 365–370.

Schomerus, G., & Angermeyer, M. C. (2008). Stigma and its impact on help-seeking for mental disorders: what do we know? *Epidemiologia E Psichiatria Sociale*, *17*(01), 31–37.

Schünemann, H. J., Oxman, A. D., Vist, G. E., Higgins, J. P., Deeks, J. J., Glasziou, P., & Guyatt, G. H. (2008). Interpreting results and drawing conclusions. *Cochrane Handbook for Systematic Reviews of Interventions. Chichester: Wiley-Blackwell*, 359–87.

Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The Role of Numeracy in Understanding the Benefit of Screening Mammography. *Annals of Internal Medicine*, *127*(11), 966–972.

Schwartz, L. M., Woloshin, S., Dvorin, E. L., & Welch, H. G. (2006). Ratio measures in leading medical journals: structured review of accessibility

of underlying absolute risks. *BMJ*, *333*(7581), 1248–. doi:10.1136/bmj.38985.564317.7C

Sheeran, P., & Abraham, C. (1996). The health belief model. In M. Connor & P. Norman, *Predicting health behaviour* (Vol. 2, pp. 29–80).

Skogstad, P., Deane, F. P., & Spicer, J. (2006). Social-cognitive determinants of help-seeking for mental health problems among prison inmates. *Criminal Behaviour and Mental Health*, *16*(1), 43–59. doi:10.1002/cbm.54

Slovic, P., Peters, E., Finucane, M. L., & MacGregor, D. G. (2005). Affect, risk, and decision making. *Health Psychology*, *24*(4S), S35.

Smeeth, L., Haines, A., & Ebrahim, S. (1999). Numbers needed to treat derived from meta-analyses—sometimes informative, usually misleading. *BMJ*, *318*(7197), 1548–1551. doi:10.1136/bmj.318.7197.1548

Smith, J. P., Tran, G. Q., & Thompson, R. D. (2008). Can the theory of planned behavior help explain men's psychological help-seeking? Evidence for a mediation effect and clinical implications. *Psychology of Men & Masculinity*, *9*(3), 179.

Smith, V. K., & Desvousges, W. H. (1987). An Empirical Analysis of the Economic Value of Risk Changes. *The Journal of Political Economy*, *95*(1), 89–114.

Sniehotta, F. F., Presseau, J., & Araújo-Soares, V. (2014). Time to retire the theory of planned behaviour. *Health Psychology Review*, *8*(1), 1–7. doi:10.1080/17437199.2013.869710

Spielmans, G. I., Berman, M. I., & Usitalo, A. N. (2011). Psychotherapy versus second-generation antidepressants in the treatment of depression: a

meta-analysis. *The Journal of Nervous and Mental Disease*, *199*(3), 142–149. doi:10.1097/NMD.0b013e31820caefb

Spitzer, R. L., Kroenke, K., Williams, J. B. ., & Lowe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, *166*(10), 1092.

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, *94*(4), 672–695. doi:10.1037/0022-3514.94.4.672

Statistics, O. for N. (2003). *Ethnic Group Statistics: A Guide for the Collection and Classification of Ethnicity Data*. The Stationery Office London.

Stecker, T., Fortney, J., Hamilton, F., Sherbourne, C. D., & Ajzen, I. (2010). Engagement in mental health treatment among veterans returning from Iraq. *Patient Preference and Adherence*, *4*, 45.

Stone, E. R., Yates, J. F., & Parker, A. M. (1997). Effects of numerical and graphical displays on professed risk-taking behavior. *Journal of Experimental Psychology: Applied*, *3*(4), 243.

Stovring, H., Gyrd-Hansen, D., Kristiansen, I., Nexoe, J., & Nielsen, J. (2008). Communicating effectiveness of intervention for chronic diseases: what single format can replace comprehensive information? *BMC Medical Informatics and Decision Making*, *8*(1), 25.

Swanson, W. H., & Cohen, J. M. (2003). Color vision. *Ophthalmology Clinics of North America*, *16*(2), 179–203.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (Vol. 4). London: Allyn and Bacon.

Tuckett, D. (1976). Becoming a patient. In D. Tuckett (Ed.), *An introduction to medical sociology.* (pp. 169–189). London: Tavistock.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124–1131. doi:10.1126/science.185.4157.1124

Vogel, D. L., Wester, S. R., Wei, M., & Boysen, G. A. (2005). The Role of Outcome Expectations and Attitudes on Decisions to Seek Professional Help. *Journal of Counseling Psychology*, *52*(4), 459.

Vogt, F., Hall, S., Hankins, M., & Marteau, T. (2009). Evaluating three theory-based interventions to increase physicians' recommendations of smoking cessation services. *Health Psychology*, *28*(2), 174–182.

Vogt, F., Mason, D., & Marteau, T. M. (2012). Crediting treatments for good outcomes that would have happened anyway: The impact of baseline risk on treatment perceptions. *Medical Decision Making*. doi:10/15/0272989X11419858

Wahl, O. F. (1997). *Media Madness: Public Images of Mental Illness*. Rutgers University Press.

Walker, A., Grimshaw, J., Johnston, M., Pitts, N., Steen, N., & Eccles, M. (2003). PRIME - PRocess modelling in ImpleMEntation research: selecting a theoretical basis for interventions to change clinical practice. *BMC Health Services Research*, *3*(1), 22.

Webb, T. L., & Sheeran, P. (2006). Does Changing Behavioral Intentions Engender Behavior Change? A Meta-Analysis of the Experimental Evidence. *Psychological Bulletin*, *132*(2), 249–268. doi:10.1037/0033-2909.132.2.249

Weinstein, N. D., Kwitel, A., McCaul, K. D., Magnan, R. E., Gerrard, M., & Gibbons, F. X. (2007). Risk Perceptions: Assessment and Relationship to Influenza Vaccination. *Health Psychology*, *26*(2), 146–151.

Weisberg, H. I., Hayden, V. C., & Pontes, V. P. (2009). Selection criteria and generalizability within the counterfactual framework: explaining the paradox of antidepressant-induced suicidality? *Clinical Trials*, *6*(2), 109–118.

Westerhof, G. J., Maessen, M., de Bruijn, R., & Smets, B. (2008). Intentions to seek (preventive) psychological help among older adults: An application of the theory of planned behaviour. *Aging & Mental Health*, *12*(3), 317–322. doi:10.1080/13607860802120797

WHO. (2014). Metrics: Disability-Adjusted Life Year (DALY). *Work Health Organisation*. Retrieved 24 April 2014, from URL:http://www.who.int/healthinfo/global_burden_disease/metrics_daly/en/. Accessed: 2014-04-24. (Archived by WebCite® at http://www.webcitation.org/6P4hQQHOa)

Wilson, D. B. (2001). Practical Meta-Analysis Effect Size Calculator (Version http://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-OR1.php). Retrieved from http://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-OR1.php

Wishart, G. C., Azzato, E. M., Greenberg, D. C., Rashbass, J., Kearins, O., Lawrence, G., … Pharoah, P. D. (2010). PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. *Breast Cancer Research*, *12*(1), R1. doi:10.1186/bcr2464

Wright, K. B. (2005). Researching Internet-Based Populations: Advantages and Disadvantages of Online Survey Research, Online Questionnaire Authoring Software Packages, and Web Survey Services. *Journal of Computer-Mediated Communication*, *10*(3), 00–00. doi:10.1111/j.1083-6101.2005.tb00259.x

Yeh, C. J. (2002). Taiwanese students' gender, age, interdependent and independent self-construal, and collective self-esteem as predictors of professional psychological help-seeking attitudes. *Cultural Diversity and Ethnic Minority Psychology*, *8*(1), 19.

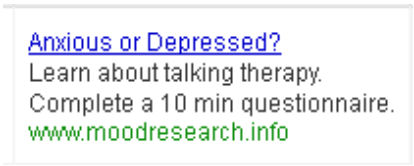
Zikmund-Fisher, B. J., Fagerlin, A., & Ubel, P. A. (2010). A Demonstration of '"Less Can Be More"' in Risk Graphics. *Medical Decision Making*, *30*(6), 661–671. doi:10.1177/0272989X10364244

Zikmund-Fisher, B. J., Smith, D. M., Ubel, P. A., & Fagerlin, A. (2007). Validation of the Subjective Numeracy Scale: Effects of Low Numeracy on Comprehension of Risk Communications and Utility Elicitations. *Medical Decision Making*, *27*(5), 663–671. doi:10.1177/0272989x07303824

Zikmund-Fisher, B. J., Witteman, H. O., Dickson, M., Fuhrel-Forbis, A., Kahn, V. C., Exe, N. L., … Fagerlin, A. (2013). Blocks, Ovals, or People? Icon Type Affects Risk Perceptions and Recall of Pictographs. *Medical Decision Making*, 0272989X13511706. doi:10.1177/0272989X13511706

Appendixes

Appendix 1: Google advertisement



Appendix 2: Google search terms

treatment for depression
anxiety treatment
clinical depression treatment
depression treatment
post traumatic stress disorder
social anxiety
anxiety depression
treatment for anxiety
PTSD
treatment of depression
depression symptoms
bipolar depression
natural treatment for depression
depression treatments
depression help
anxiety attack treatment
treatment for anxiety attacks
symptoms of panic attacks
GAD
anxiety problems
depression and anxiety treatment
treatment depression
treatments for depression
anxiety treatments
how to treat anxiety attacks
anxiety depression treatment
postpartum depression treatment
manic depression treatment
treatment resistant depression
depression treatment options
help with anxiety
dealing with depression and anxiety
anxiety and depression treatment
treatments for anxiety
treatment for depression and anxiety
treating anxiety attacks

major depression treatment
severe depression treatment
anxiety attack relief
stop anxiety attacks
how to cure anxiety attacks
cure anxiety attacks
dealing with anxiety attacks
ocd symptoms
stress and anxiety symptoms
anxiety attacks treatment
cure for anxiety attacks
low mood
treatment for anxiety and depression
how to control anxiety attacks
best treatment for depression
cure anxiety and panic attacks
help with depression and anxiety

Appendix 3: Email documenting RHUL ethics approval

Ref: 2013/043 Ethics Form Approved

Psychology-Webmaster@rhul.ac.uk

Fri 17/05/2013 17:55

To:nwjt093@rhul.ac.uk <nwjt093@rhul.ac.uk>; Wroe, Abigail <Abigail.Wroe@rhul.ac.uk>;

Cc:PSY-EthicsAdmin@rhul.ac.uk <PSY-EthicsAdmin@rhul.ac.uk>; Leman, Patrick <Patrick.Leman@rhul.ac.uk>;

Application Details:

Applicant Name: **Florian Vogt**

Application title: **Decisions about whether or not to opt for psychological therapy**

Comments: Approved, but please note some comments from one of the reviewers which might help you to clarify some of the associated documentation.
Best,
Patrick

Reviewers' comments for consideration.
I have a few minor comments:

Section C3: Will participants be asked whether they have direct experience of talking therapies or particular CMHPs themselves? Will any exclusion criteria be used?

Section C5: The following part of this section isn't clear: 'with the exception of being able to collect data from participants without prior notification'. What is meant by 'without prior notification'? Additionally, it is stated: 'Approval will be sought from the relevant ethics committee'. Will approval be sought from additional ethics committees?

Questionnaire, question 34: Typo: 'educational' should be 'education'.

RHUL Information Sheet, paragraph 2: 'How effective are talking-therapy?' should be 'How effective is talking-therapy?'

Appendix 4: Questionnaire

Department of Psychology
Royal Holloway, University of London
Egham, Surrey TW20 0EX, UK

Information Sheet

Investigating beliefs about the benefit of talking-therapy for mental health problems

This is a study on beliefs about the benefit of talking-therapy conducted by Dr Florian Vogt, Trainee Clinical Psychologist and supervised by Dr Abigail Wroe, Lecturer in Psychology, both from Royal Holloway, University of London.

We would appreciate your participation, because the benefits of treatments are often misinterpreted. This study aims to increase our understanding of how information about the treatment benefits is interpreted.

- If you decide to take part, I will ask you to read asked to read information about talking-therapy for mental health problems.
- You will then complete some questions about it, for example: 'How effective are talking-therapy?'.
- This will take around 10 minutes and is conducted entirely over the internet.
- Anyone aged 18 and above can take part in this study.

Nobody except us will be allowed to see the data you provide. In this study you will be known only by number; so the information you provide is completely anonymous. You do not have to take part in this study if you don't want to. Not taking part in this study will not affect you in any way. If you decide to take part you are still free to withdraw at any moment without having to give a reason; as participation is anonymous it is only possible to withdraw before you submit the questionnaire. The anonymity means that we cannot identify your questionnaire once submission has occurred. Please be aware that submission of the questionnaire implies consent.

If you would like to discuss any aspect of the research with Dr Abigail Wroe, you can contact her by email (Abigail.Wroe@rhul.ac.uk) or by phone on 01784 276532. If you would like to contact me, please email (Florian.Vogt.2011@live.rhul.ac.uk) or call 01784 443851.

Please keep this sheet yourself for reference. Please feel free to ask any questions before you complete the questionnaire form below. This study has been reviewed and approved by the Psychology Department internal ethical procedure at Royal Holloway, University of London.

1. Consent form

Have you (please tick all that apply):

- [ ] Read the information sheet about the study?
- [ ] Understood that the study is completely anonymous?
- [ ] Understood you're free to withdraw from the study until you submit the questionnaire, without giving a reason (and without it affecting your care/education if applicable)?
- [ ] Do you agree to take part in the study?

Beliefs about talking-therapy

Page description:
Negative thoughts and feelings such as anxiety, feeling low, or low self-esteem are very common.

Talking-therapy or psychological therapy involves talking to someone who is trained to help people deal with negative thoughts and feelings and supports them to make positive changes. There are different kinds of talking-therapy and they include:

- counselling
- psychotherapy
- cognitive behavioural therapy (CBT)

In order to answer the following questions please imagine that you experience at least some negative thoughts and feelings, if you do not.

2. By using talking-therapy, I will have fewer bothersome symptoms (get better).

	1	2	3	4	5	6	7	
Extremely Unlikely	○	○	○	○	○	○	○	**Extremely Likely**

3. If I use talking-therapy, people will think I am crazy.

	1	2	3	4	5	6	7	
Extremely Unlikely	○	○	○	○	○	○	○	**Extremely Likely**

4. Using talking-therapy means that I am failing (i.e. cannot handle problems) on my own.

	1	2	3	4	5	6	7	
Extremely Unlikely	○	○	○	○	○	○	○	**Extremely Likely**

5. Using talking-therapy will hurt my career.

	1	2	3	4	5	6	7	
Extremely Unlikely	○	○	○	○	○	○	○	**Extremely Likely**

6. Using talking-therapy will help me identify my triggers and learn to cope better.

	1	2	3	4	5	6	7	
Extremely Unlikely	○	○	○	○	○	○	○	**Extremely Likely**

7. For me to identify my triggers and learn to cope better is ...

	1	2	3	4	5	6	7	
Extremely Undesirable	○	○	○	○	○	○	○	**Extremely Desirable**

8. For me to have fewer bothersome symptoms (get better) is ...

	1	2	3	4	5	6	7	
Extremely Undesirable	○	○	○	○	○	○	○	**Extremely Desirable**

9. Failing (i.e. not handling problems) on my own would be ...

	1	2	3	4	5	6	7	
Extremely Undesirable	○	○	○	○	○	○	○	**Extremely Desirable**

10. If people thought I was crazy, this would be ...

	1	2	3	4	5	6	7	
Extremely Undesirable	○	○	○	○	○	○	○	**Extremely Desirable**

11. If my work career was hurt, this would be ...

	1	2	3	4	5	6	7	
Extremely Undesirable	○	○	○	○	○	○	○	**Extremely Desirable**

12. Overall, I think using talking-therapy for me is ...

	1	2	3	4	5	6	7	
Worthless	○	○	○	○	○	○	○	**Valuable**

13. Overall, I think using talking-therapy for me is ...

	1	2	3	4	5	6	7	
Harmful	○	○	○	○	○	○	○	**Beneficial**

14. Overall, I think using talking-therapy for me is ...

	1	2	3	4	5	6	7	
Not Therapeutic	○	○	○	○	○	○	○	**Therapeutic**

15. Overall, I think using talking-therapy for me is ...

	1	2	3	4	5	6	7	
Bad	○	○	○	○	○	○	○	**Good**

16. Most people who are important to me would approve of my using talking-therapy.

	1	2	3	4	5	6	7	
Definitely False	○	○	○	○	○	○	○	**Definitely True**

17. It is expected of me that I use talking-therapy if I have emotional and behavioural problems.

	1	2	3	4	5	6	7	
Definitely False	○	○	○	○	○	○	○	**Definitely True**

18. People who are close to me themselves use talking-therapy if they have emotional and behavioural problems.

	1	2	3	4	5	6	7	
Definitely False	○	○	○	○	○	○	○	**Definitely True**

19. If I wanted, I could easily use talking-therapy.

	1	2	3	4	5	6	7	
Definitely False	○	○	○	○	○	○	○	**Definitely True**

20. Using talking-therapy is under my control.

	1	2	3	4	5	6	7	
Definitely False	○	○	○	○	○	○	○	**Definitely True**

21. I am confident that I could use talking-therapy if I wanted to.

	1	2	3	4	5	6	7	
Definitely False	○	○	○	○	○	○	○	**Definitely True**

22. I intend to use talking-therapy for emotional and behavioural problems.

	1	2	3	4	5	6	7	
Definitely False	○	○	○	○	○	○	○	**Definitely True**

23. I will try to use talking-therapy for emotional and behavioural problems.

	1	2	3	4	5	6	7	
Extremely Unlikely	○	○	○	○	○	○	○	**Extremely Likely**

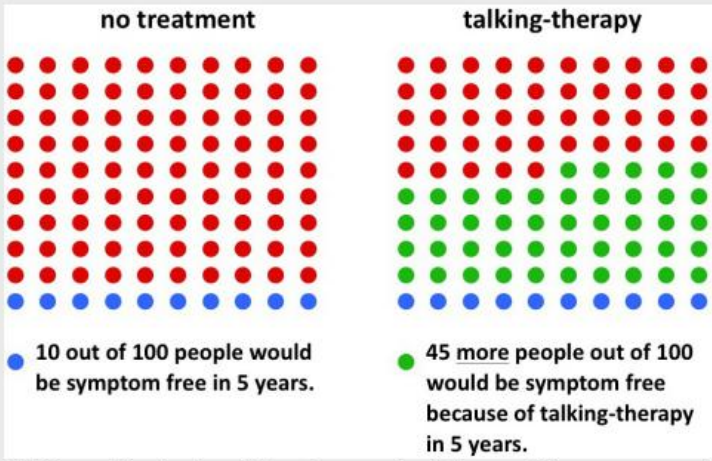
24. I plan to use talking-therapy for emotional and behavioural problems.

	1	2	3	4	5	6	7	
Strongly Disagree	○	○	○	○	○	○	○	**Strongly Agree**

Judging effectiveness of talking-therapy

Page description:
The information shown about the effectiveness of talking-therapy on this page is fictional.

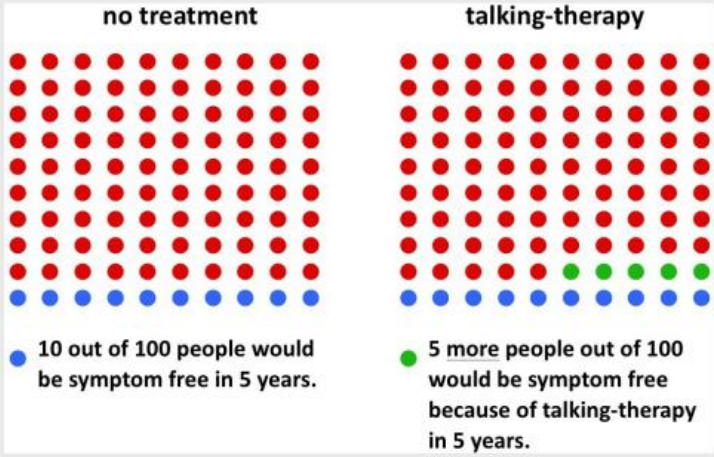
Imagine that you have been diagnosed with Obsessive Compulsive Disorder (OCD). Treatment includes talking-therapy. In the graphic below you can see the outcomes for people who choose no treament and for people who choose talking-therapy.



25. How effective is talking-therapy for **this condition** compared to no treatment? *

ineffective — extremely effective

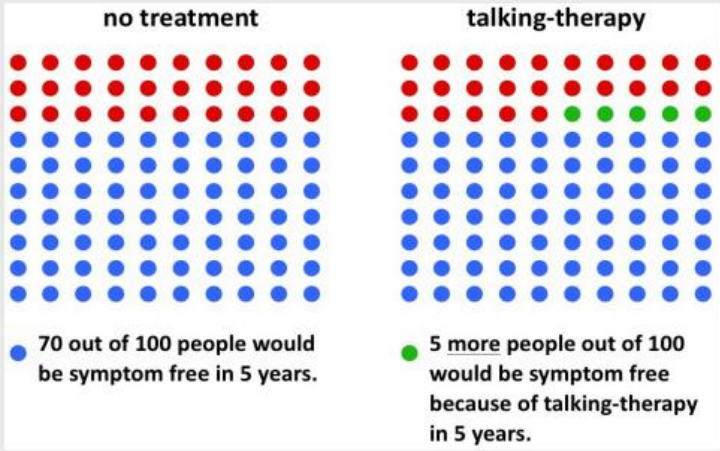
Imagine that you have been diagnosed with Depression. Treatment includes talking-therapy. In the graphic below you can see the outcomes for people who choose no treament and for people who choose talking-therapy.



26. How effective is talking-therapy for **this condition** compared to no treatment? *

ineffective ——— extremely effective

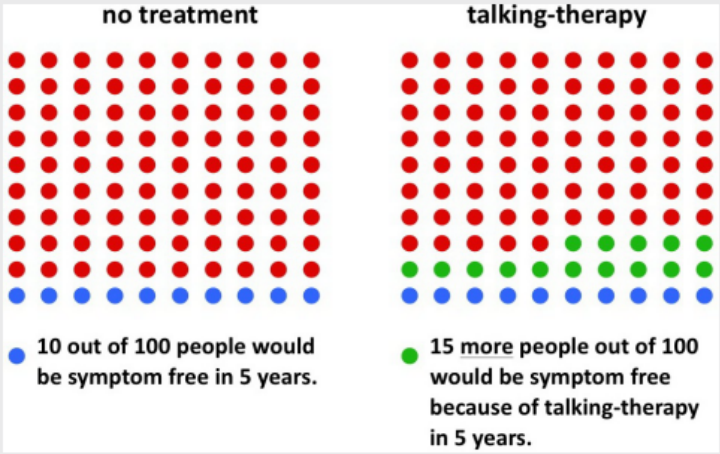
Imagine that you have been diagnosed with Chronic Low Self-Esteem. Treatment includes talking-therapy. In the graphic below you can see the outcomes for people who choose no treament and for people who choose talking-therapy.



27. How effective is talking-therapy for **this condition** compared to no treatment? *

ineffective — extremely effective

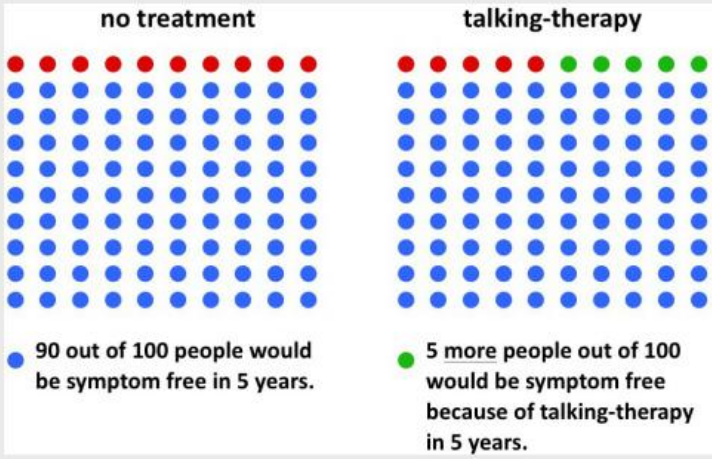
Imagine that you have been diagnosed with Post-Traumatic Stress Disorder (PTSD). Treatment includes talking-therapy. In the graphic below you can see the outcomes for people who choose no treament and for people who choose talking-therapy.



28. How effective is talking-therapy for **this condition** compared to no treatment? *

ineffective — extremely effective

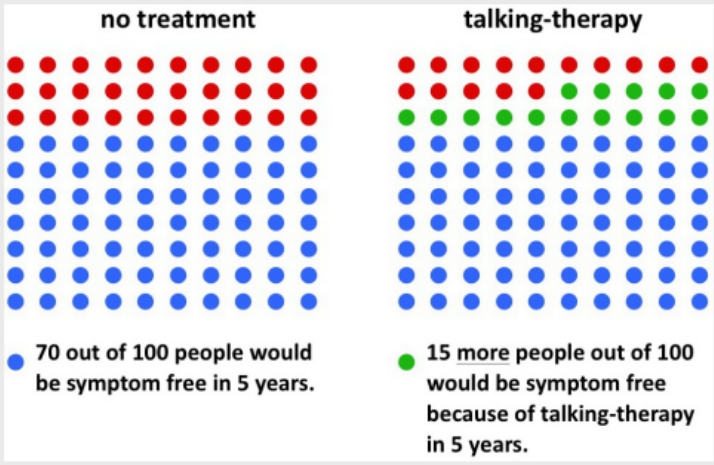
Imagine that you have been diagnosed with Generalised Anxiety Disorder (GAD). Treatment includes talking-therapy. In the graphic below you can see the outcomes for people who choose no treament and for people who choose talking-therapy.



29. How effective is talking-therapy for **this condition** compared to no treatment? *

ineffective — extremely effective

Imagine that you have been diagnosed with Social Anxiety Disorder. Treatment includes talking-therapy. In the graphic below you can see the outcomes for people who choose no treament and for people who choose talking-therapy.



30. How effective is talking-therapy for **this condition** compared to no treatment? *

ineffective — extremely effective

Information about you

Page description:

31. I am.

- Male
- Female
- Other

32. How old are you?

33. Are you currently receiving pharmacological or talking-therapy treatment for a mental health problem?

- Yes
- No

34. Have you ever received pharmacological or talking-therapy treatment for a mental health problem?

- Yes
- No

35. Over the last 2 weeks, how often have you been bothered by any of the following problems?

	Not at all	Several days	More than half the days	Nearly every day
Little interest or pleasure in doing things				
Feeling down, depressed, or hopeless				
Feeling nervous, anxious or on edge				
Not being able to stop or control worrying				

36. What is your highest level of education completed?

- ○ No education completed
- ○ Primary (primary school)
- ○ Secondary 1 (compulsory education, GCSE, O level, or similar)
- ○ Secondary 2 (admission to university, A level or similar)
- ○ Tertiary (university and other forms of education after the secondary level)

37. How would you describe your ethnic group?

- ○ Asian
- ○ Black
- ○ Chinese
- ○ Mixed
- ○ White

Number skills

Page description:

38. Imagine that we flip a fair coin 1,000 times. What is your best guess about how many times the coin would come up heads in 1,000 flips? *

... times out of 1,000

39. In the BIG CASH LOTTERY, the chance of winning a £10 prize is 1%. What is your best guess about how many people would win a £10 prize if 1000 people each buy a single ticket for BIG CASH LOTTERY? *

... person(s) out of 1,000

40. In the BARGAIN CARS SWEEPSTAKES, the chance of winning a car is 1 in 1,000. What percent of tickets from BARGAIN CARS SWEEPSTAKES win a car? *

... %

Thank You!

Debrief Information

Thank you for participating in the study. Below is a description of the background, purpose and potential implications of this research project.

Many people with mental illness do not use talking therapy. This is a source of concern because effective treatments are available. There are many reasons why adults do not seek professional help; one of these is not believing that talking therapy can help.

When people judge whether a treatment is effective they are often influenced by the rate at which people get better with treatment. But for many health conditions, such as the flu or other infections, people would get better even if they had not used treatment. The actual effect treatment is often much smaller than assumed.

However, most patients with a mental health condition remain unwell without treatment. In effect that means that even though treatment for mental condition helps many people, some need several treatment attempts.

This study investigated whether talking-therapy are judged as unduly ineffective because the natural rate of getting better can be very low.

Here is some information on the actual treatment effectiveness:

Generalised anxiety disorder (GAD)
86% of patients with generalised anxiety disorder (GAD) remain unwell without treatment. **Talking-therapy helps 32% of patients**. It means that 54% remain unwell even though the treatment is very effective.

Post-Traumatic-Stress-Disorder (PTSD)
89% of patients with Post-Traumatic-Stress-Disorder (PTSD) remain unwell without treatment. **Talking-therapy helps 48% of patients.** It means that 41% remain unwell even though the treatment is very effective.

Depression
72% of patients with Depression remain unwell without treatment. **Talking-therapy helps 19% of patients.** It means that 54% remain unwell even though the treatment is very effective.

Further information and contact details

If you would like to receive a copy of the final report or have any questions or comments, please email (Florian.Vogt.2011@live.rhul.ac.uk).

Appendix 5: SPSS output of analysis

Hyp 1

Within-Subjects Factors

Measure:MEASURE_1

RRR	ARR	Dependent Variable
1	1	v70plus5
	2	v90plus5
2	1	v10plus15
	2	v70plus15

Between-Subjects Factors

		Value Label	N
Numeracy	.00		50
	1.00		71
	2.00		51
	3.00		29
MEDIAN(PHQ4,ALL)	1	1	45
(Binned)	2	2	156
MEDIAN(gender,ALL)	1.0		51
	2.0		150

Tests of Within-Subjects Contrasts

Measure:MEASURE_1

Source	ARR	Type I Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
RRR	Linear	7967.613	1	7967.613	29.666	.000	.134	29.666	1.000
RRR * NumeracyScale	Linear	788.904	3	262.968	.979	.404	.015	2.937	.264
RRR * PHQ4_1_L_H	Linear	12.768	1	12.768	.048	.828	.000	.048	.055
RRR * ctreatment_1	Linear	136.170	1	136.170	.507	.477	.003	.507	.109
RRR * ptreatment_1	Linear	13.028	1	13.028	.049	.826	.000	.049	.056

RRR * gender_1	Linear		497.098	1	497.098	1.851	.175	.010	1.851	.273
RRR * age_1	Linear		461.307	1	461.307	1.718	.192	.009	1.718	.256
RRR * edu_1	Linear		170.616	1	170.616	.635	.426	.003	.635	.125
Error(RRR)	Linear		51298.745	191	268.580					
ARR		Linear	1464.210	1	1464.210	5.363	.022	.027	5.363	.635
ARR * NumeracyScale		Linear	610.324	3	203.441	.745	.526	.012	2.235	.208
ARR * PHQ4_1_L_H		Linear	100.920	1	100.920	.370	.544	.002	.370	.093
ARR * ctreatment_1		Linear	123.542	1	123.542	.452	.502	.002	.452	.103
ARR * ptreatment_1		Linear	33.033	1	33.033	.121	.728	.001	.121	.064
ARR * gender_1		Linear	32.560	1	32.560	.119	.730	.001	.119	.064
ARR * age_1		Linear	16.281	1	16.281	.060	.807	.000	.060	.057
ARR * edu_1		Linear	3.446	1	3.446	.013	.911	.000	.013	.051
Error(ARR)		Linear	52148.934	191	273.031					
RRR * ARR	Linear	Linear	1246.270	1	1246.270	5.292	.023	.027	5.292	.629
RRR * ARR * NumeracyScale	Linear	Linear	977.494	3	325.831	1.384	.249	.021	4.151	.364
RRR * ARR * PHQ4_1_L_H	Linear	Linear	19.318	1	19.318	.082	.775	.000	.082	.059
RRR * ARR * ctreatment_1	Linear	Linear	.014	1	.014	.000	.994	.000	.000	.050
RRR * ARR * ptreatment_1	Linear	Linear	18.865	1	18.865	.080	.777	.000	.080	.059
RRR * ARR * gender_1	Linear	Linear	56.067	1	56.067	.238	.626	.001	.238	.077
RRR * ARR * age_1	Linear	Linear	13.453	1	13.453	.057	.811	.000	.057	.057
RRR * ARR * edu_1	Linear	Linear	15.001	1	15.001	.064	.801	.000	.064	.057

Error(RRR*ARR)	Linear	Linear	44980.767	191	235.501				

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
30-5 (70 plus 5)	.957	14	186	.499
10-5 (90 plus 5)	1.560	14	186	.094
90-15 (10 plus 15)	1.058	14	186	.399
30-15 (70 plus 15)	1.417	14	186	.149

Tests of Between-Subjects Effects

Measure:MEASURE_1

Transformed Variable:Average

Source	Type I Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	1878136.668	1	1878136.668	706.848	.000	.787	706.848	1.000
NumeracyScale	23299.396	3	7766.465	2.923	.035	.044	8.769	.689
PHQ4_1_L_H	1764.397	1	1764.397	.664	.416	.003	.664	.128
ctreatment_1	2052.610	1	2052.610	.773	.381	.004	.773	.141
ptreatment_1	1670.450	1	1670.450	.629	.429	.003	.629	.124
gender_1	10951.730	1	10951.730	4.122	.044	.021	4.122	.524
age_1	885.981	1	885.981	.333	.564	.002	.333	.089
edu_1	9074.837	1	9074.837	3.415	.066	.018	3.415	.452
Error	507498.181	191	2657.059					

Estimated Marginal Means

1. Grand Mean

Measure:MEASURE_1

Mean	Std. Error	95% Confidence Interval	
		Lower Bound	Upper Bound
51.385[a]	2.495	46.464	56.306

2. RRR * ARR

Measure:MEASURE_1

RRR	ARR	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound

1	1	48.008[a]	2.869	42.349	53.667
	2	49.421[a]	2.986	43.532	55.310
2	1	51.397[a]	2.713	46.045	56.748
	2	56.714[a]	2.770	51.249	62.178

3. RRR

Estimates

Measure:MEASURE_1

			95% Confidence Interval	
RRR	Mean	Std. Error	Lower Bound	Upper Bound
1	48.714[a]	2.733	43.323	54.106
2	54.055[a]	2.497	49.129	58.981

Pairwise Comparisons

Measure:MEASURE_1

					95% Confidence Interval for Difference[a]	
(I) RRR	(J) RRR	Mean Difference (I-J)	Std. Error	Sig.[a]	Lower Bound	Upper Bound
1	2	-5.341[*]	1.586	.001	-8.470	-2.211
2	1	5.341[*]	1.586	.001	2.211	8.470

Multivariate Tests

	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Pillai's trace	.056	11.333[a]	1.000	191.000	.001	.056	11.333	.918
Wilks' lambda	.944	11.333[a]	1.000	191.000	.001	.056	11.333	.918
Hotelling's trace	.059	11.333[a]	1.000	191.000	.001	.056	11.333	.918
Roy's largest root	.059	11.333[a]	1.000	191.000	.001	.056	11.333	.918

4. ARR

Estimates

Measure:MEASURE_1

			95% Confidence Interval	
ARR	Mean	Std. Error	Lower Bound	Upper Bound

1	49.702[a]	2.537	44.698	54.707
2	53.067[a]	2.700	47.741	58.393

Pairwise Comparisons

Measure:MEASURE_1

(I) ARR	(J) ARR	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1	2	-3.365*	1.600	.037	-6.520	-.210
2	1	3.365*	1.600	.037	.210	6.520

Multivariate Tests

	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Pillai's trace	.023	4.425[a]	1.000	191.000	.037	.023	4.425	.553
Wilks' lambda	.977	4.425[a]	1.000	191.000	.037	.023	4.425	.553
Hotelling's trace	.023	4.425[a]	1.000	191.000	.037	.023	4.425	.553
Roy's largest root	.023	4.425[a]	1.000	191.000	.037	.023	4.425	.553

5. Numeracy

Estimates

Measure:MEASURE_1

Numeracy	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
.00	58.666[a]	4.312	50.160	67.171
1.00	46.853[a]	3.670	39.615	54.091
2.00	54.191[a]	3.883	46.532	61.850
3.00	45.830[a]	5.069	35.832	55.827

Pairwise Comparisons

Measure:MEASURE_1

(I) Numeracy	(J) Numeracy	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound

.00	1.00	11.813*	4.880	.016	2.187	21.439
	2.00	4.475	5.523	.419	-6.419	15.368
	3.00	12.836*	6.417	.047	.179	25.493
1.00	.00	-11.813*	4.880	.016	-21.439	-2.187
	2.00	-7.338	4.930	.138	-17.063	2.387
	3.00	1.023	5.884	.862	-10.583	12.629
2.00	.00	-4.475	5.523	.419	-15.368	6.419
	1.00	7.338	4.930	.138	-2.387	17.063
	3.00	8.361	6.108	.173	-3.686	20.409
3.00	.00	-12.836*	6.417	.047	-25.493	-.179
	1.00	-1.023	5.884	.862	-12.629	10.583
	2.00	-8.361	6.108	.173	-20.409	3.686

Univariate Tests

Measure:MEASURE_1

	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Contrast	5110.402	3	1703.467	2.564	.056	.039	7.693	.625
Error	126874.545	191	664.265					

6. MEDIAN(gender,ALL)

Estimates

Measure:MEASURE_1

			95% Confidence Interval	
MEDIAN(gender,ALL)	Mean	Std. Error	Lower Bound	Upper Bound
1.0	54.863[a]	3.958	47.056	62.670
2.0	47.907[a]	2.601	42.775	53.038

Pairwise Comparisons

Measure:MEASURE_1

(I) MEDIAN(gender,ALL)	(J) MEDIAN(gender,ALL)	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1.0	2.0	6.956	4.469	.121	-1.858	15.770
2.0	1.0	-6.956	4.469	.121	-15.770	1.858

Univariate Tests

Measure:MEASURE_1

	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Contrast	1609.742	1	1609.742	2.423	.121	.013	2.423	.341
Error	126874.545	191	664.265					

Hyp 2

Within-Subjects Factors

Measure:MEASURE_1

RRR	CGO	Dependent Variable
1	1	v10plus15
	2	v70plus5
2	1	v10plus45
	2	v70plus15

Between-Subjects Factors

		Value Label	N
Numeracy	.00		50
	1.00		71
	2.00		51
	3.00		29
MEDIAN(PHQ4,ALL)	1	1	45
(Binned)	2	2	156
MEDIAN(gender,ALL)	1.0		51
	2.0		150

Tests of Within-Subjects Contrasts

Measure:MEASURE_1

Source	CGO	Type I Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
RRR	Linear	31732.091	1	31732.091	165.462	.000	.464	165.462	1.000
RRR * NumeracyScale	Linear	3305.692	3	1101.897	5.746	.001	.083	17.237	.947

RRR * PHQ4_1_L_H	Linear		285.497	1	285.497	1.489	.224	.008	1.489	.229
RRR * ctreatment_1	Linear		323.462	1	323.462	1.687	.196	.009	1.687	.253
RRR * ptreatment_1	Linear		300.642	1	300.642	1.568	.212	.008	1.568	.238
RRR * gender_1	Linear		1690.627	1	1690.627	8.816	.003	.044	8.816	.840
RRR * age_1	Linear		3.831	1	3.831	.020	.888	.000	.020	.052
RRR * edu_1	Linear		667.714	1	667.714	3.482	.064	.018	3.482	.459
Error(RRR)	Linear		36629.695	191	191.779					
CGO		Linear	10934.359	1	10934.359	25.495	.000	.118	25.495	.999
CGO * NumeracyScale		Linear	3473.731	3	1157.910	2.700	.047	.041	8.099	.650
CGO * PHQ4_1_L_H		Linear	126.472	1	126.472	.295	.588	.002	.295	.084
CGO * ctreatment_1		Linear	839.954	1	839.954	1.958	.163	.010	1.958	.286
CGO * ptreatment_1		Linear	253.925	1	253.925	.592	.443	.003	.592	.119
CGO * gender_1		Linear	779.711	1	779.711	1.818	.179	.009	1.818	.269
CGO * age_1		Linear	5.403	1	5.403	.013	.911	.000	.013	.051
CGO * edu_1		Linear	567.547	1	567.547	1.323	.251	.007	1.323	.208
Error(CGO)		Linear	81918.148	191	428.891					
RRR * CGO	Linear	Linear	2561.225	1	2561.225	15.737	.000	.076	15.737	.977
RRR * CGO * NumeracyScale	Linear	Linear	697.420	3	232.473	1.428	.236	.022	4.285	.375
RRR * CGO * PHQ4_1_L_H	Linear	Linear	10.743	1	10.743	.066	.798	.000	.066	.058
RRR * CGO * ctreatment_1	Linear	Linear	303.834	1	303.834	1.867	.173	.010	1.867	.275
RRR * CGO * ptreatment_1	Linear	Linear	63.714	1	63.714	.391	.532	.002	.391	.095

RRR * CGO * gender_1	Linear	Linear	172.013	1	172.013	1.057	.305	.006	1.057	.176
RRR * CGO * age_1	Linear	Linear	239.807	1	239.807	1.473	.226	.008	1.473	.227
RRR * CGO * edu_1	Linear	Linear	214.166	1	214.166	1.316	.253	.007	1.316	.207
Error(RRR*CGO)	Linear	Linear	31085.327	191	162.750					

Tests of Between-Subjects Effects

Measure:MEASURE_1

Transformed Variable:Average

Source	Type I Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	2281070.429	1	2281070.429	966.823	.000	.835	966.823	1.000
NumeracyScale	15619.885	3	5206.628	2.207	.089	.034	6.620	.554
PHQ4_1_L_H	2386.859	1	2386.859	1.012	.316	.005	1.012	.170
ctreatment_1	262.626	1	262.626	.111	.739	.001	.111	.063
ptreatment_1	857.291	1	857.291	.363	.547	.002	.363	.092
gender_1	4794.098	1	4794.098	2.032	.156	.011	2.032	.294
age_1	1278.619	1	1278.619	.542	.463	.003	.542	.113
edu_1	4565.101	1	4565.101	1.935	.166	.010	1.935	.283
Error	450635.342	191	2359.347					

Estimated Marginal Means

1. Grand Mean

Measure:MEASURE_1

Mean	Std. Error	95% Confidence Interval	
		Lower Bound	Upper Bound
55.838[a]	2.351	51.201	60.476

2. CGO * RRR

Measure:MEASURE_1

CGO	RRR	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1	1	51.397[a]	2.713	46.045	56.748
	2	67.235[a]	2.487	62.330	72.140

2	1	48.008[a]	2.869	42.349	53.667
	2	56.714[a]	2.770	51.249	62.178

3. Numeracy * RRR

Measure:MEASURE_1

				95% Confidence Interval	
Numeracy	RRR	Mean	Std. Error	Lower Bound	Upper Bound
.00	1	57.349[a]	4.385	48.699	65.999
	2	63.711[a]	4.059	55.705	71.716
1.00	1	46.284[a]	3.732	38.922	53.645
	2	56.113[a]	3.454	49.300	62.926
2.00	1	51.689[a]	3.949	43.899	59.478
	2	67.222[a]	3.655	60.013	74.432
3.00	1	43.488[a]	5.155	33.320	53.656
	2	60.852[a]	4.771	51.441	70.262

4. Numeracy * CGO

Measure:MEASURE_1

				95% Confidence Interval	
Numeracy	CGO	Mean	Std. Error	Lower Bound	Upper Bound
.00	1	60.734[a]	4.209	52.431	69.037
	2	60.326[a]	4.616	51.222	69.430
1.00	1	55.743[a]	3.582	48.677	62.809
	2	46.654[a]	3.928	38.906	54.402
2.00	1	62.299[a]	3.791	54.822	69.776
	2	56.612[a]	4.156	48.413	64.810
3.00	1	58.487[a]	4.948	48.727	68.248
	2	45.852[a]	5.426	35.150	56.554

5. RRR * MEDIAN(gender,ALL)

Measure:MEASURE_1

				95% Confidence Interval	
RRR	MEDIAN(gender,ALL)	Mean	Std. Error	Lower Bound	Upper Bound
1	1.0	53.425[a]	4.025	45.485	61.364
	2.0	45.980[a]	2.646	40.761	51.199
2	1.0	62.665[a]	3.725	55.317	70.013
	2.0	61.283[a]	2.449	56.454	66.113

Hyp 3

Within-Subjects Factors

Measure:MEASURE_1

ARR	CGO	Dependent Variable
1	1	v10plus5
	2	v70plus5
2	1	v10plus15
	2	v70plus15

Between-Subjects Factors

		Value Label	N
Numeracy	.00		50
	1.00		71
	2.00		51
	3.00		29
MEDIAN(PHQ4,ALL)	1	1	45
(Binned)	2	2	156
MEDIAN(gender,ALL)	1.0		51
	2.0		150

Tests of Within-Subjects Contrasts

Measure:MEASURE_1

Source	CGO	Type I Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
ARR	Linear	12317.922	1	12317.922	75.838	.000	.284	75.838	1.000
ARR * NumeracyScale	Linear	818.121	3	272.707	1.679	.173	.026	5.037	.435
ARR * PHQ4_1_L_H	Linear	126.226	1	126.226	.777	.379	.004	.777	.142
ARR * ctreatment_1	Linear	80.063	1	80.063	.493	.483	.003	.493	.108
ARR * ptreatment_1	Linear	200.853	1	200.853	1.237	.268	.006	1.237	.198
ARR * gender_1	Linear	434.060	1	434.060	2.672	.104	.014	2.672	.370
ARR * age_1	Linear	188.789	1	188.789	1.162	.282	.006	1.162	.189

ARR * edu_1	Linear		169.794	1	169.794	1.045	.308	.005	1.045	.174
Error(ARR)	Linear		31022.923	191	162.424					
CGO		Linear	3252.101	1	3252.101	8.629	.004	.043	8.629	.832
CGO * NumeracyScale		Linear	1710.149	3	570.050	1.513	.213	.023	4.538	.395
CGO * PHQ4_1_L_H		Linear	10.671	1	10.671	.028	.867	.000	.028	.053
CGO * ctreatment_1		Linear	6.777	1	6.777	.018	.893	.000	.018	.052
CGO * ptreatment_1		Linear	38.680	1	38.680	.103	.749	.001	.103	.062
CGO * gender_1		Linear	36.315	1	36.315	.096	.757	.001	.096	.061
CGO * age_1		Linear	16.566	1	16.566	.044	.834	.000	.044	.055
CGO * edu_1		Linear	14.758	1	14.758	.039	.843	.000	.039	.054
Error(CGO)		Linear	71983.732	191	376.878					
ARR * CGO	Linear	Linear	273.583	1	273.583	1.858	.174	.010	1.858	.274
ARR * CGO * NumeracyScale	Linear	Linear	470.088	3	156.696	1.064	.365	.016	3.193	.285
ARR * CGO * PHQ4_1_L_H	Linear	Linear	5.683	1	5.683	.039	.844	.000	.039	.054
ARR * CGO * ctreatment_1	Linear	Linear	70.452	1	70.452	.479	.490	.002	.479	.106
ARR * CGO * ptreatment_1	Linear	Linear	23.188	1	23.188	.158	.692	.001	.158	.068
ARR * CGO * gender_1	Linear	Linear	51.376	1	51.376	.349	.555	.002	.349	.090
ARR * CGO * age_1	Linear	Linear	13.712	1	13.712	.093	.761	.000	.093	.061
ARR * CGO * edu_1	Linear	Linear	3.330	1	3.330	.023	.881	.000	.023	.053
Error(ARR*CGO)	Linear	Linear	28118.338	191	147.216					

Tests of Between-Subjects Effects

Measure:MEASURE_1

Transformed Variable:Average

Source	Type I Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	1819063.494	1	1819063.494	685.652	.000	.782	685.652	1.000
NumeracyScale	20532.860	3	6844.287	2.580	.055	.039	7.739	.628
PHQ4_1_L_H	1179.432	1	1179.432	.445	.506	.002	.445	.102
ctreatment_1	1813.426	1	1813.426	.684	.409	.004	.684	.130
ptreatment_1	2645.448	1	2645.448	.997	.319	.005	.997	.168
gender_1	10647.952	1	10647.952	4.013	.047	.021	4.013	.513
age_1	1406.509	1	1406.509	.530	.467	.003	.530	.112
edu_1	9068.836	1	9068.836	3.418	.066	.018	3.418	.452
Error	506730.793	191	2653.041					

Estimated Marginal Means

1. Grand Mean

Measure:MEASURE_1

Mean	Std. Error	95% Confidence Interval	
		Lower Bound	Upper Bound
50.209[a]	2.493	45.291	55.126

2. ARR * CGO

Measure:MEASURE_1

ARR	CGO	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1	1	44.717[a]	2.833	39.129	50.306
	2	48.008[a]	2.869	42.349	53.667
2	1	51.397[a]	2.713	46.045	56.748
	2	56.714[a]	2.770	51.249	62.178

3. MEDIAN(gender,ALL)

Estimates

Measure:MEASURE_1

MEDIAN(gender,ALL)	Mean	Std. Error	95% Confidence Interval

			Lower Bound	Upper Bound
1.0	53.599[a]	3.955	45.798	61.400
2.0	46.818[a]	2.600	41.691	51.946

Pairwise Comparisons

Measure:MEASURE_1

(I) MEDIAN(gender,ALL)	(J) MEDIAN(gender,ALL)	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1.0	2.0	6.781	4.465	.131	-2.027	15.588
2.0	1.0	-6.781	4.465	.131	-15.588	2.027

Univariate Tests

Measure:MEASURE_1

	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Contrast	1529.587	1	1529.587	2.306	.131	.012	2.306	.327
Error	126682.698	191	663.260					

Appendix 6: Alternative analysis with PHQ2 and GAD2

Hyp 1 (phq/gad)

Within-Subjects Factors

Measure:MEASURE_1

RRR	ARR	Dependent Variable
1	1	v70plus5
	2	v90plus5
2	1	v10plus15
	2	v70plus15

Between-Subjects Factors

		N
Numeracy	.00	45
	1.00	68
	2.00	46
	3.00	27
MEDIAN(gender,ALL)	1.0	49
	2.0	137

Tests of Within-Subjects Contrasts

Measure:MEASURE_1

Source	ARR	Type I Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
RRR	Linear	6907.647	1	6907.647	25.463	.000	.127	25.463	.999
RRR * NumeracyScale	Linear	609.260	3	203.087	.749	.525	.013	2.246	.208
RRR * ctreatment_1	Linear	120.614	1	120.614	.445	.506	.003	.445	.102
RRR * ptreatment_1	Linear	12.052	1	12.052	.044	.833	.000	.044	.055
RRR * gender_1	Linear	433.560	1	433.560	1.598	.208	.009	1.598	.242
RRR * age_1	Linear	455.879	1	455.879	1.680	.197	.010	1.680	.252

RRR * edu_1	Linear		151.758	1	151.758	.559	.456	.003	.559	.115
RRR * PHQ2	Linear		11.574	1	11.574	.043	.837	.000	.043	.055
RRR * GAD2	Linear		78.207	1	78.207	.288	.592	.002	.288	.083
Error(RRR)	Linear		47475.199	175	271.287					
ARR		Linear	1373.819	1	1373.819	5.190	.024	.029	5.190	.620
ARR * NumeracyScale		Linear	594.646	3	198.215	.749	.524	.013	2.247	.208
ARR * ctreatment_1		Linear	77.020	1	77.020	.291	.590	.002	.291	.084
ARR * ptreatment_1		Linear	31.331	1	31.331	.118	.731	.001	.118	.064
ARR * gender_1		Linear	40.029	1	40.029	.151	.698	.001	.151	.067
ARR * age_1		Linear	53.505	1	53.505	.202	.654	.001	.202	.073
ARR * edu_1		Linear	90.049	1	90.049	.340	.560	.002	.340	.089
ARR * PHQ2		Linear	200.998	1	200.998	.759	.385	.004	.759	.139
ARR * GAD2		Linear	171.278	1	171.278	.647	.422	.004	.647	.126
Error(ARR)		Linear	46320.076	175	264.686					
RRR * ARR	Linear	Linear	1170.012	1	1170.012	5.770	.017	.032	5.770	.666
RRR * ARR * NumeracyScale	Linear	Linear	1232.017	3	410.672	2.025	.112	.034	6.075	.514
RRR * ARR * ctreatment_1	Linear	Linear	14.606	1	14.606	.072	.789	.000	.072	.058
RRR * ARR * ptreatment_1	Linear	Linear	299.276	1	299.276	1.476	.226	.008	1.476	.227
RRR * ARR * gender_1	Linear	Linear	89.091	1	89.091	.439	.508	.003	.439	.101
RRR * ARR * age_1	Linear	Linear	64.756	1	64.756	.319	.573	.002	.319	.087
RRR * ARR * edu_1	Linear	Linear	5.753	1	5.753	.028	.866	.000	.028	.053

RRR * ARR * PHQ2	Linear	Linear	11.516	1	11.516	.057	.812	.000	.057	.056
RRR * ARR * GAD2	Linear	Linear	.008	1	.008	.000	.995	.000	.000	.050
Error(RRR*ARR)	Linear	Linear	35487.714	175	202.787					

Tests of Between-Subjects Effects

Measure:MEASURE_1

Transformed Variable:Average

Source	Type I Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	1754053.259	1	1754053.259	655.499	.000	.789	655.499	1.000
NumeracyScale	19603.108	3	6534.369	2.442	.066	.040	7.326	.601
ctreatment_1	1301.472	1	1301.472	.486	.486	.003	.486	.107
ptreatment_1	2018.471	1	2018.471	.754	.386	.004	.754	.139
gender_1	9251.961	1	9251.961	3.458	.065	.019	3.458	.456
age_1	167.031	1	167.031	.062	.803	.000	.062	.057
edu_1	11196.364	1	11196.364	4.184	.042	.023	4.184	.530
PHQ2	4243.202	1	4243.202	1.586	.210	.009	1.586	.240
GAD2	1499.216	1	1499.216	.560	.455	.003	.560	.115
Error	468283.664	175	2675.907					

Estimated Marginal Means

1. Grand Mean

Measure:MEASURE_1

Mean	Std. Error	95% Confidence Interval	
		Lower Bound	Upper Bound
50.271[a]	2.233	45.864	54.679

2. RRR * ARR

Measure:MEASURE_1

RRR	ARR	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1	1	47.276[a]	2.539	42.264	52.287
	2	48.118[a]	2.662	42.865	53.372
2	1	50.487[a]	2.428	45.695	55.279

2. RRR * ARR

Measure:MEASURE_1

RRR	ARR	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1	1	47.276[a]	2.539	42.264	52.287
	2	48.118[a]	2.662	42.865	53.372
2	1	50.487[a]	2.428	45.695	55.279
	2	55.204[a]	2.456	50.357	60.051

3. RRR

Estimates

Measure:MEASURE_1

RRR	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	47.697[a]	2.452	42.857	52.536
2	52.846[a]	2.230	48.444	57.247

Pairwise Comparisons

Measure:MEASURE_1

(I) RRR	(J) RRR	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1	2	-5.149[*]	1.422	.000	-7.955	-2.342
2	1	5.149[*]	1.422	.000	2.342	7.955

Multivariate Tests

	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Pillai's trace	.070	13.106[a]	1.000	175.000	.000	.070	13.106	.950
Wilks' lambda	.930	13.106[a]	1.000	175.000	.000	.070	13.106	.950
Hotelling's trace	.075	13.106[a]	1.000	175.000	.000	.070	13.106	.950
Roy's largest root	.075	13.106[a]	1.000	175.000	.000	.070	13.106	.950

4. ARR

Estimates

Measure:MEASURE_1

ARR	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	48.881[a]	2.252	44.438	53.325
2	51.661[a]	2.427	46.870	56.452

Pairwise Comparisons

Measure:MEASURE_1

(I) ARR	(J) ARR	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1	2	-2.780[*]	1.405	.049	-5.552	-.007
2	1	2.780[*]	1.405	.049	.007	5.552

Multivariate Tests

	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Pillai's trace	.022	3.916[a]	1.000	175.000	.049	.022	3.916	.503
Wilks' lambda	.978	3.916[a]	1.000	175.000	.049	.022	3.916	.503
Hotelling's trace	.022	3.916[a]	1.000	175.000	.049	.022	3.916	.503
Roy's largest root	.022	3.916[a]	1.000	175.000	.049	.022	3.916	.503

5. Numeracy

Estimates

Measure:MEASURE_1

Numeracy	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
.00	57.083[a]	4.442	48.317	65.849
1.00	45.355[a]	3.408	38.629	52.081
2.00	53.048[a]	3.954	45.244	60.852
3.00	45.599[a]	5.057	35.619	55.578

Pairwise Comparisons

Measure:MEASURE_1

(I) Numeracy	(J) Numeracy	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
.00	1.00	11.728*	5.093	.022	1.675	21.780
	2.00	4.034	5.854	.492	-7.519	15.588
	3.00	11.484	6.729	.090	-1.797	24.765
1.00	.00	-11.728*	5.093	.022	-21.780	-1.675
	2.00	-7.693	5.141	.136	-17.839	2.453
	3.00	-.244	6.121	.968	-12.325	11.837
2.00	.00	-4.034	5.854	.492	-15.588	7.519
	1.00	7.693	5.141	.136	-2.453	17.839
	3.00	7.450	6.402	.246	-5.186	20.085
3.00	.00	-11.484	6.729	.090	-24.765	1.797
	1.00	.244	6.121	.968	-11.837	12.325
	2.00	-7.450	6.402	.246	-20.085	5.186

Univariate Tests

Measure:MEASURE_1

	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Contrast	4482.597	3	1494.199	2.234	.086	.037	6.701	.558
Error	117070.916	175	668.977					

6. MEDIAN(gender,ALL)

Estimates

Measure:MEASURE_1

MEDIAN(gender,ALL)	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1.0	53.601[a]	3.874	45.956	61.246
2.0	46.942[a]	2.415	42.176	51.707

Pairwise Comparisons

Measure:MEASURE_1

(I) MEDIAN(gender,AL	(J) MEDIAN(gender,AL	Mean Difference	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]

L)	L)	(I-J)			Lower Bound	Upper Bound
1.0	2.0	6.659	4.661	.155	-2.539	15.858
2.0	1.0	-6.659	4.661	.155	-15.858	2.539

Univariate Tests

Measure:MEASURE_1

	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Contrast	1365.733	1	1365.733	2.042	.155	.012	2.042	.295
Error	117070.916	175	668.977					

Hyp 2 (phq/gad)

Within-Subjects Factors

Measure:MEASURE_1

RRR	CGO	Dependent Variable
1	1	v10plus15
	2	v70plus5
2	1	v10plus45
	2	v70plus15

Between-Subjects Factors

		N
Numeracy	.00	45
	1.00	68
	2.00	46
	3.00	27
MEDIAN(gender,ALL)	1.0	49
	2.0	137

Tests of Within-Subjects Contrasts

Measure:MEASURE_1

Source	CGO	Type I Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]

RRR	Linear		27825.941	1	27825.941	149.541	.000	.461	149.541	1.000
RRR * NumeracyScale	Linear		1801.453	3	600.484	3.227	.024	.052	9.681	.735
RRR * ctreatment_1	Linear		377.613	1	377.613	2.029	.156	.011	2.029	.294
RRR * ptreatment_1	Linear		157.688	1	157.688	.847	.359	.005	.847	.150
RRR * gender_1	Linear		1389.481	1	1389.481	7.467	.007	.041	7.467	.776
RRR * age_1	Linear		71.889	1	71.889	.386	.535	.002	.386	.095
RRR * edu_1	Linear		1018.998	1	1018.998	5.476	.020	.030	5.476	.643
RRR * PHQ2	Linear		310.221	1	310.221	1.667	.198	.009	1.667	.250
RRR * GAD2	Linear		46.011	1	46.011	.247	.620	.001	.247	.078
Error(RRR)	Linear		32563.206	175	186.075					
CGO		Linear	9128.005	1	9128.005	21.014	.000	.107	21.014	.995
CGO * NumeracyScale		Linear	2947.502	3	982.501	2.262	.083	.037	6.786	.564
CGO * ctreatment_1		Linear	594.704	1	594.704	1.369	.244	.008	1.369	.214
CGO * ptreatment_1		Linear	588.527	1	588.527	1.355	.246	.008	1.355	.212
CGO * gender_1		Linear	462.682	1	462.682	1.065	.303	.006	1.065	.177
CGO * age_1		Linear	47.377	1	47.377	.109	.742	.001	.109	.062
CGO * edu_1		Linear	401.351	1	401.351	.924	.338	.005	.924	.159
CGO * PHQ2		Linear	46.638	1	46.638	.107	.744	.001	.107	.062
CGO * GAD2		Linear	40.844	1	40.844	.094	.759	.001	.094	.061
Error(CGO)		Linear	76015.870	175	434.376					
RRR * CGO	Linear	Linear	2174.710	1	2174.710	13.094	.000	.070	13.094	.949

RRR * CGO * NumeracyScale	Linear	Linear	680.324	3	226.775	1.365	.255	.023	4.096	.359
RRR * CGO * ctreatment_1	Linear	Linear	296.733	1	296.733	1.787	.183	.010	1.787	.265
RRR * CGO * ptreatment_1	Linear	Linear	12.169	1	12.169	.073	.787	.000	.073	.058
RRR * CGO * gender_1	Linear	Linear	102.551	1	102.551	.617	.433	.004	.617	.122
RRR * CGO * age_1	Linear	Linear	407.526	1	407.526	2.454	.119	.014	2.454	.344
RRR * CGO * edu_1	Linear	Linear	102.280	1	102.280	.616	.434	.004	.616	.122
RRR * CGO * PHQ2	Linear	Linear	.001	1	.001	.000	.998	.000	.000	.050
RRR * CGO * GAD2	Linear	Linear	229.451	1	229.451	1.381	.241	.008	1.381	.215
Error(RRR*CGO)	Linear	Linear	29065.756	175	166.090					

Tests of Between-Subjects Effects

Measure:MEASURE_1

Transformed Variable:Average

Source	Type I Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	2114560.344	1	2114560.344	908.274	.000	.838	908.274	1.000
NumeracyScale	13151.265	3	4383.755	1.883	.134	.031	5.649	.482
ctreatment_1	61.898	1	61.898	.027	.871	.000	.027	.053
ptreatment_1	1441.524	1	1441.524	.619	.432	.004	.619	.123
gender_1	4256.030	1	4256.030	1.828	.178	.010	1.828	.270
age_1	198.475	1	198.475	.085	.771	.000	.085	.060
edu_1	6952.314	1	6952.314	2.986	.086	.017	2.986	.405
PHQ2	4702.610	1	4702.610	2.020	.157	.011	2.020	.293
GAD2	2027.149	1	2027.149	.871	.352	.005	.871	.153
Error	407418.891	175	2328.108					

Estimated Marginal Means

1. Grand Mean

Measure:MEASURE_1

Mean	Std. Error	95% Confidence Interval	
		Lower Bound	Upper Bound
54.597[a]	2.083	50.486	58.709

2. CGO * RRR

Measure:MEASURE_1

CGO	RRR	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1	1	50.487[a]	2.428	45.695	55.279
	2	65.423[a]	2.202	61.077	69.769
2	1	47.276[a]	2.539	42.264	52.287
	2	55.204[a]	2.456	50.357	60.051

Hyp 3 (phq/gad)

Within-Subjects Factors

Measure:MEASURE_1

ARR	CGO	Dependent Variable
1	1	v10plus5
	2	v70plus5
2	1	v10plus15
	2	v70plus15

Between-Subjects Factors

		N
Numeracy	.00	45
	1.00	68
	2.00	46
	3.00	27
MEDIAN(gender,ALL)	1.0	49
	2.0	137

Tests of Within-Subjects Contrasts

Measure:MEASURE_1

Source	CGO	Type I Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]

ARR	Linear		10795.102	1	10795.102	67.654	.000	.279	67.654	1.000
ARR * NumeracyScale	Linear		407.732	3	135.911	.852	.467	.014	2.555	.233
ARR * ctreatment_1	Linear		76.521	1	76.521	.480	.490	.003	.480	.106
ARR * ptreatment_1	Linear		273.286	1	273.286	1.713	.192	.010	1.713	.256
ARR * gender_1	Linear		321.595	1	321.595	2.015	.157	.011	2.015	.292
ARR * age_1	Linear		353.041	1	353.041	2.213	.139	.012	2.213	.316
ARR * edu_1	Linear		355.222	1	355.222	2.226	.137	.013	2.226	.317
ARR * PHQ2	Linear		95.279	1	95.279	.597	.441	.003	.597	.120
ARR * GAD2	Linear		.608	1	.608	.004	.951	.000	.004	.050
Error(ARR)	Linear		27923.613	175	159.564					
CGO		Linear	3024.194	1	3024.194	8.042	.005	.044	8.042	.805
CGO * NumeracyScale		Linear	1616.191	3	538.730	1.433	.235	.024	4.298	.375
CGO * ctreatment_1		Linear	2.519	1	2.519	.007	.935	.000	.007	.051
CGO * ptreatment_1		Linear	17.976	1	17.976	.048	.827	.000	.048	.055
CGO * gender_1		Linear	42.900	1	42.900	.114	.736	.001	.114	.063
CGO * age_1		Linear	30.087	1	30.087	.080	.778	.000	.080	.059
CGO * edu_1		Linear	79.690	1	79.690	.212	.646	.001	.212	.074
CGO * PHQ2		Linear	8.794	1	8.794	.023	.879	.000	.023	.053
CGO * GAD2		Linear	90.946	1	90.946	.242	.623	.001	.242	.078
Error(CGO)		Linear	65810.704	175	376.061					
ARR * CGO	Linear	Linear	264.968	1	264.968	1.852	.175	.010	1.852	.273

ARR * CGO * NumeracyScale	Linear	Linear	337.026	3	112.342	.785	.504	.013	2.355	.217
ARR * CGO * ctreatment_1	Linear	Linear	42.789	1	42.789	.299	.585	.002	.299	.085
ARR * CGO * ptreatment_1	Linear	Linear	55.686	1	55.686	.389	.534	.002	.389	.095
ARR * CGO * gender_1	Linear	Linear	84.933	1	84.933	.594	.442	.003	.594	.119
ARR * CGO * age_1	Linear	Linear	97.548	1	97.548	.682	.410	.004	.682	.130
ARR * CGO * edu_1	Linear	Linear	8.768	1	8.768	.061	.805	.000	.061	.057
ARR * CGO * PHQ2	Linear	Linear	61.127	1	61.127	.427	.514	.002	.427	.100
ARR * CGO * GAD2	Linear	Linear	515.774	1	515.774	3.604	.059	.020	3.604	.471
Error(ARR*CGO)	Linear	Linear	25041.381	175	143.094					

Tests of Between-Subjects Effects

Measure:MEASURE_1

Transformed Variable:Average

Source	Type I Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	1699423.876	1	1699423.876	644.381	.000	.786	644.381	1.000
NumeracyScale	15783.002	3	5261.001	1.995	.117	.033	5.985	.507
ctreatment_1	1145.223	1	1145.223	.434	.511	.002	.434	.100
ptreatment_1	3362.508	1	3362.508	1.275	.260	.007	1.275	.202
gender_1	8704.532	1	8704.532	3.301	.071	.019	3.301	.439
age_1	239.816	1	239.816	.091	.763	.001	.091	.060
edu_1	12620.549	1	12620.549	4.785	.030	.027	4.785	.585
PHQ2	3455.192	1	3455.192	1.310	.254	.007	1.310	.207
GAD2	846.599	1	846.599	.321	.572	.002	.321	.087
Error	461526.702	175	2637.295					

Estimated Marginal Means

1. Grand Mean

Measure:MEASURE_1

Mean	Std. Error	95% Confidence Interval	
		Lower Bound	Upper Bound
49.231[a]	2.217	44.856	53.607

2. ARR * CGO

Measure:MEASURE_1

ARR	CGO	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1	1	43.959[a]	2.520	38.986	48.932
	2	47.276[a]	2.539	42.264	52.287
2	1	50.487[a]	2.428	45.695	55.279
	2	55.204[a]	2.456	50.357	60.051

3. MEDIAN(gender,ALL)

Estimates

Measure:MEASURE_1

MEDIAN(gender,ALL)	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1.0	52.350[a]	3.846	44.760	59.940
2.0	46.113[a]	2.397	41.382	50.844

Pairwise Comparisons

Measure:MEASURE_1

(I) MEDIAN(gender,ALL)	(J) MEDIAN(gender,ALL)	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1.0	2.0	6.236	4.627	.179	-2.896	15.368
2.0	1.0	-6.236	4.627	.179	-15.368	2.896

Univariate Tests

Measure:MEASURE_1

	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Contrast	1197.727	1	1197.727	1.817	.179	.010	1.817	.268
Error	115381.676	175	659.324					