Impact of a Progressive Stepped Care Approach in an Improving Access to Psychological Therapies Service: An Observational Study

Boyd, Lisa Susan

Impact of a Progressive Stepped Care Approach in an Improving Access to Psychological Therapies Service: An Observational Study

Volume 1/2

Lisa Susan Boyd

Thesis submitted for the degree of Master of Science in Research in The School of Medicine and Health at Durham University.

February 2016

Statement of Copyright

Acknowledgments

Professor Joe Reilly for patience, support and supervision.

Dr David Ekers for having faith in me, and being very persuasive in encouraging me to undertake a research project, and giving ongoing support.

Louise Mainwaring for initial support regarding data extraction.

Emma Baker, a huge thank you for technical and statistical support. It has been a huge learning curve for us both, and I could not have completed this without Emma's support.

Contents

Tables

Figures

Appendices

Chapter 1: Introduction

Mental health problems such as depression and anxiety disorders are common, with a general understanding that one in six people in their lifetime are likely to suffer with a mental health problem at some point, and can be particularly debilitating affecting employment, relationships and general quality of life (1). The cost to society is difficult to estimate, however health economic analyses describes how investing in treating mental illness with therapy pays for itself by an increase in people working, and reducing the cost to the welfare state. (2,3) The evidence base for efficacy of psychological interventions in common mental disorders such as depression and anxiety is well-established, however less is known about how to implement them in local service contexts. In addition, the notion that severe and complex should require a more intensive treatment perpetuates service model design and clinical guidance, yet there is a growing body of evidence of the efficacy of low intensity interventions. The Stepped Care model is a framework of organising a range of treatment and intensity and is recommended by The National Institute for Health Care Excellence. (NICE). However, there appears limited evidence regarding the effectiveness of a stepped care model, different interpretations of what stepped care actually means in routine practice, and even more uncertainty as to whether outcomes are affected by one type of stepped care model compared to others.

The Improving Access to Psychological Therapies (IAPT) programme in England aims to implement psychological therapies for common mental disorders such as depression and anxiety on a large scale, and broadly takes a stepped care approach. Its goals are major clinical benefits, but also economic benefits in terms of return to work. The collation of standardised minimum dataset collection of information is mandatory, providing information such as patient demographics, categorisation of referred problem, psychological severity measures administered every treatment session, type of treatment delivered, number of treatment sessions and use of psychological measure scores to measure clinical outcomes such as recovery. Such detailed information across many services nationally provides for the first time a comprehensive picture of people with common mental health problems,

treatments delivered, and the possibility of exploring what works, particularly in routine practice service delivery. The service delivery model of therapies appears to be heterogeneous across the country, particularly in terms of different versions of stepped care. The study site is an IAPT service which has implemented a range of initiatives to improve outcome, one of which is a move from a stratified, allocated model of stepped care, to a progression model of delivery. This service development has created a naturalistic opportunity for measuring how this service change is associated with outcomes, particularly for those scoring higher on psychological measures, indicating severity of problem. This project aim was to explore the impact on clinical outcomes moving from one service delivery model (allocated) to another (progression) and therefore there were two main research questions for this study:

1. What is the relationship between clinical outcomes of depression and anxiety and service delivery model for adults treated in and IAPT service?

2. What is the relationship between the clinical outcomes for moderate to severe anxiety and depression, and service delivery model?

Baseline variables such as gender, ethnicity, disability and employment status were also analysed to explore any relationship with outcome and model.

Chapter 2: Background & Context

There is a substantial body of research regarding the efficacy of effectiveness of psychological interventions, arguably the largest being in the modality of Cognitive behavioural therapy (CBT), particularly with regard to common mental health problems, anxiety and depressive disorders, which has led to several clinical guidelines. (4-11) In particular with over forty years of research on psychotherapy for depression, it is of note that a recent network meta-analysis comparing seven different psychotherapeutic approaches found that there was no significant difference in effectiveness between any of the types of therapy (12). Despite an extensive body of research which informs national guidelines and influences how mental health treatment is delivered, there remains a lack of evidence regarding what specifically works for whom. Psychotherapy trials are often underpowered to demonstrate clear findings, and there is no clear superiority between modality of therapies and clinical outcomes (12, 13). In addition such questions relating to the efficacy of specific interventions, little evidence exists to guide clinician and managers on regarding service design and delivery, and how this may impact on clinical outcome.

Traditionally therapeutic intervention has been informed by professional training and latest research, emphasis on types of therapy on offer varied locally, with service design being much more locally defined within mental health and psychology services. The design and introduction of the national programme in England of Improving Access to Psychological Therapies (IAPT) was a comprehensive attempt to increase the number of professionally qualified therapists nationally, thereby increasing the availability of talking therapies, through a structured training programme and new service design. Its predicted benefits are the reduction of the cost of prescribed medication, and the reduction of the state welfare cost with more people previously unemployed or on sickness benefit, able to return to work through an improvement or recovery of their mental health. (11, 14)

2.1 Definition of stepped care

Of historical importance is a review and discussion paper (MAPLE) by Lovell & Richards (15) predating the recommendation of the stepped care model by NICE; in many ways was an early proposal of a version of stepped care. The authors outline the problems with traditional Cognitive Behavioural Therapy delivery, and propose a solution which is multi access points and levels of entry for CBT services (MAPLE) as a service delivery model with what looks like an early version of stepped care. It is of particular importance to highlight the issues the authors identified with traditional therapy delivery which were time wasted through clients not attending, these sessions are wasted. In terms of dosage, session length is traditionally 50-60 minutes planned around therapist convenience and traditional, however there no evidence of optimal session time. The authors also discuss a "negatively accelerated dose affect curve", which questions whether the traditional number of sessions offered (NICE 12-20 sessions dependant on disorder) is actually needed by all or majority of patients. The "efficacy of brief vs intensive therapy" is also explored, and the argument is outlined that brief and low intensity interventions are effective for a larger patient group than is traditionally treated, and therefore a system which favours these as a first line treatment, with a step up to multi strand / complex treatment if no progress at this level is recommended. (15)

A further paper (16) proposed a version of stepped care for depression, with the comparison with the stepped care treatment of physical disorders, alcohol use and the psychological treatment of some anxiety disorders. In 2005 NICE issued guidelines regarding the implementation and practice for treating depression with a stepped care approach. Pilling & Harvey (17) describe how this stepped care approach for depression draws on "chronic disease management models and the principles that the most cost-effective and least intrusive treatment should be offered first." A more stratified/allocated model of stepped care is presented, with description regarding which types of treatment should be offered depending on severity of patient presentation.

However this does not necessarily control the problems that are outlined in the MAPLE paper of the issues regarding length of time, and the "negatively accelerated dose affect curve". It is likely that given that the current NICE guidelines recommend

higher step intervention for moderate and severe presentations that this is what services and therapists would offer, even on first access to treatment, as there is the implication within this guidance that this is the correct treatment. However it still remains, that we do not know that anything up to 20 sessions of 60 minutes of a talking therapy is the best course of action.(15) In the same year as the NICE guidelines for stepped care for depression, a narrative literature review was published regarding stepped care and psychological therapies, stating that the clinical effectiveness of therapy is well evidenced but identified that there is a gap of knowledge and evidence regarding efficiencies and cost, relating to models of delivery. (18)

This review explains the definition of stepped care, with key features being described, such as 'least restrictive' – defining this as less intensive treatment, and 'self-correcting' meaning the system has mechanisms to change the treatment, i.e. 'step up' if the least restrictive treatment is not achieving results. They describe that low intensity interventions that are CBT informed e.g. problem solving and stepping up to CBT are compatible, there is a suggestion that a stepped care system could be used involving a number of different levels and types of therapy. (18)

Suggestions are made regarding how many steps, and also how decisions may be made regarding stepping up, they discuss some of the advantages and disadvantages of individual clinical decisions, and the complexity of this. As is also described in the MAPLE paper (15), "there is a general perception that minimal interventions are best restricted to less severe disorders, although the evidence for this is not definitive." The narrative review (18) considers that cost effectiveness is not a straightforward process it is also suggested that early intensive treatment may be more cost and clinically effective for complex presentations.

The narrative review (15) explores 3 key assumptions regarding issues with stepped care at the time of writing.

a) Lower intensity interventions can achieve similar clinical outcomes to that of traditional therapy for a proportion of clients.
b) Low intensity interventions enable efficiencies of resources.

c) Low intensity interventions and stepped care are acceptable to patients and clinicians.

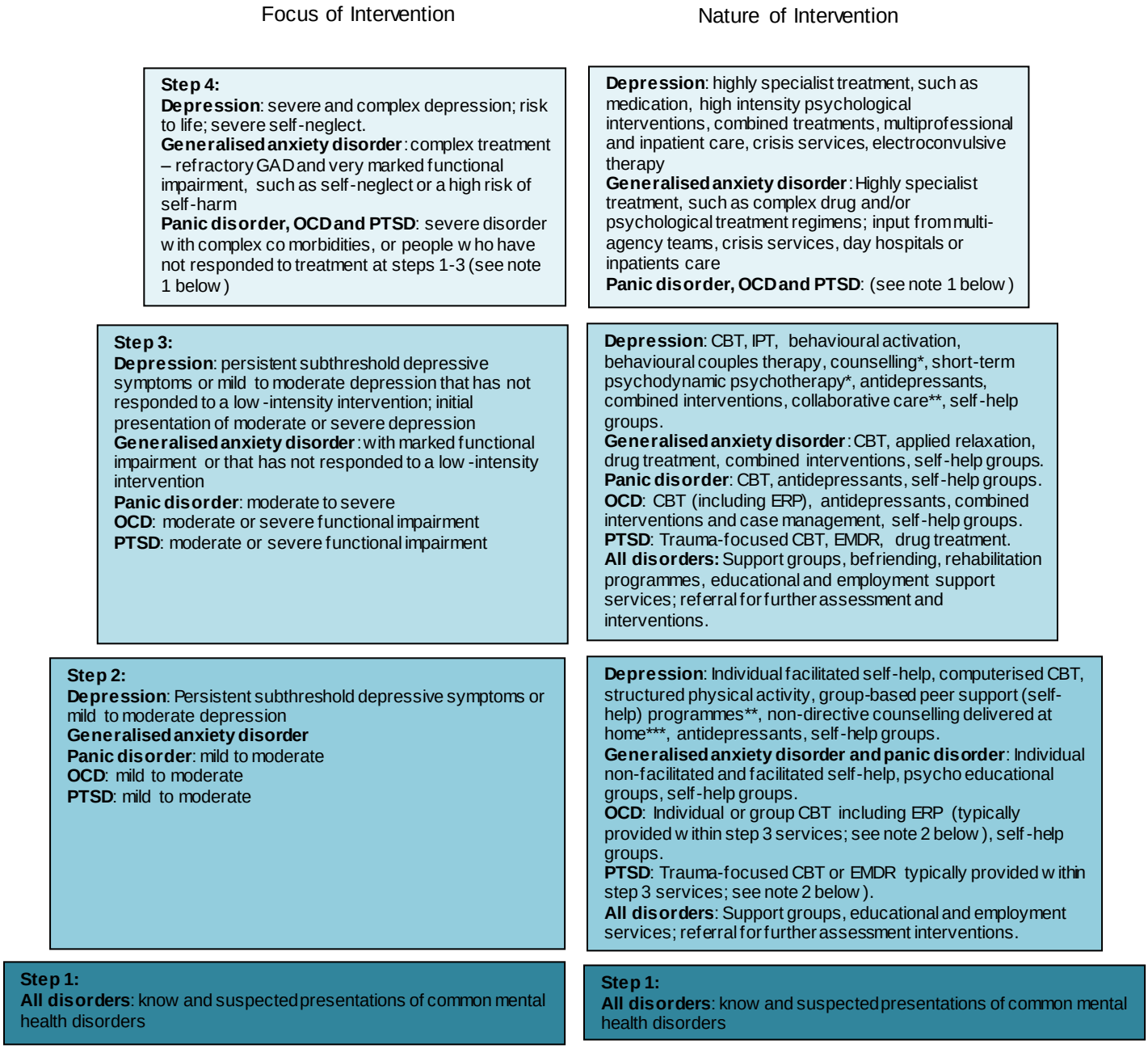
Therefore at this point although there are huge similarities with the definitions and explanations of stepped care amongst NICE guidelines and the research, there is clearly a difference in terms of interpretation of when least intrusive intervention principle is applied and how lower intensity interventions are delivered. The NICE guidelines recommended a version of stepped care that although it advocates the principle of least intrusive intervention, it appears to emphasise apparent severity of presentation influencing where in the stepped model someone begins to access treatment, therefore appearing to be more of a matched care, allocated model. This is not the same as the apparent position with evidence in the literature emerging demonstrates the validity of a more 'progressive' or pure model of stepped care, which emphasise lower intensity interventions in the first instance, even with severe presentations.

For the purpose of this study, clarity and reference here on, the definition of the different models of stepped care are as such: Patients that are allocated to an intervention based on their level of need which is predominately psychological measure and clinician judged, where least intrusive intervention may be low or high intensity, depending on presentation is defined as allocated or matched care. This is delivered by a stratified model of stepped care or could be in a mixed model where some are allocated and those that do not recover from low intensity are stepped up. A progressive or pure stepped care model is where all patients are treated with the least intrusive low intensity intervention first, and stepped up to a higher intensity intervention at the end of the first if not recovered.

Published studies have sought to demonstrate the efficacy of stepped care compared to treatment as usual, but have not compared different versions of stepped care, with the exception of two, which will be discussed alongside the context of IAPT, and the within the literature review.

2.2 Historical context to IAPT development

Mental health service community provision is generally divided into primary and secondary care, operating a stratified stepped care approach based on complexity, severity chronicity and risk. Stepped care has been a system model in operation for around 10 years, and was viewed as a potential solution to the longer standing issue of lack of numbers of qualified therapists and poor access to treatment. (16,18) Stepped care continues to be presented by NICE as the service model of delivery for psychological therapies. In the stepped care model presented in the NICE guidance ([9]) (figure 1) there are both suggestions regarding type of intervention to a particular severity of presentation, with an allocated matched care approach but also an element of progressive stepped care, to step up from lower intensity interventions if there is no improvement. Thus the current NICE guidelines appear to advocate a mixed model of delivery.



Note 1: The NICE guidance on panic disorder (CG113) and OCD (CG31) uses different models of stepped care to the step 4 model used in the NICE guidance on depression (CG90, CG91) and generalised anxiety disorder (CG113).
The NICE clinical guidance on PTSD (CG26) does not have the stepped care model. People with panic disorder, OCD or PTSD that has not responded to treatment at step 1-3, or who have severe disorders and complex co-morbidities that prevent effective management at steps 1-3, should receive specialist services at step 4, according to individual need and clinical judgement. The principle interventions at step 4 are similar to those listed for depression and generalised anxiety disorder; with the exception that electroconvulsive therapy is not indicated.
Note 2: The NICE clinical guidance on OCD (CG31) recommends that people with mild to moderate OCD receive individual or group based CBT. The NICE clinical guidance on PTSD (CG26) recommends that people with mild to moderate PTSD receive trauma-focused CBT or EMDR. These interventions may typically be commissioned from, and provided by, trained, high intensity therapy staff in step 3 services.
*Discuss with the person the uncertainty of the effectiveness of counselling and psychodynamic psychotherapy in treating depression.
** For people with depression and a chronic physical health problem.
*** For women during pregnancy or the post natal period.
Key: CBT – Cognitive behavioural therapy; ERP – exposure and response prevention; EMDR – eye movement desensitisation and reprocessing; OCD – obsessive compulsive disorder; IPT – interpersonal therapy; PTSD – post traumatic stress disorder.

Figure 1: Stepped-care model showing steps 1 to 4 for people with common mental health disorders (NICE 2011)

GP's would usually be the first point of contact prior to all steps, intervention provision from community mental health teams for steps 1-3 delivered at a primary care level, and 3-4 at a secondary care level. For patients with common mental health problems, prior to 2008, provision for talking therapies was scarce and sporadic, (19) and therefore the privilege of the more complex and chronic patient, but only with a long wait of several months. Patients with common mental health problems would be managed by GP's with prescribed medication, patients with more chronic problems managed by primary care mental health teams or secondary care services for the more risky or complex presentations. (19) However the size and cost of mental health problems is an increasing concern. "Mental ill-health accounts for over a third of all illness in Britain and 40 percent of all disability..... one in six working adults (16 percent of the population) at any one time are suffering from clinical depression and/or anxiety disorders." (19) A proposal to the government described the link between poor mental health and unemployment, and suggested that by increasing numbers of qualified therapists and targeting therapy to more people with mild to moderate mental health problems, this would increase the numbers moving into employment, thus reducing the cost to the welfare state. (3)

2.3 IAPT development

During 2006 – 2007, pilot sites tested out different service model and therapy delivery (20, 21) and produced good results regarding volume, throughput and clinical outcomes. In 2008, the 3 year roll out of a large national programme called Improving Access to Psychological Therapies (IAPT) commenced. It aimed to transform and improve the psychological therapies provision. "The Improving Access to Psychological Therapies (IAPT) programme is concerned with raising standards of recognition of, and treatment for, the mass of people who suffer from depression and anxiety disorders. The programme is at the heart of the Government's drive to give greater access to, and choice of talking therapies to those who would benefit from them." (19). IAPT was designed to be aimed at those patients that rarely accessed talking therapies, the mild to moderate range of common mental health problems, previously managed within primary care settings largely by GP's and medication, with sporadic potential involvement with a practice counsellor or the primary care mental health team. IAPT was seen as an opportunity to transform how services were delivered, and be paid for, by acting as long term prevention to more serious

mental health problems. This should also reduce demand and cost on secondary care services, and with a reduction of medication cost in the long term be more cost effective for primary care. It was also viewed as part of the solution to rising number of unemployed and the cost to the benefits system with numbers of people unemployed for long periods of time claiming mental health problems prevented them from working.

Through increasing the numbers of qualified therapists with a large scale training programme, increasing the low intensity workforce and emphasised the principle of least intrusive intervention, IAPT aims to increase the numbers accessing a talking therapy.

A comprehensive and detailed monitoring and evaluation system of individual sites and generally the IAPT programme nationally are in place. “Detailed outcome monitoring and ongoing evaluations of the programme are considered an integral part of IAPT”. (11) Each assessment and therapy session delivered is expected to collect a minimum data set (MDS), comprising of a number of psychological measures and employment and benefit status. These data are evaluated locally and nationally against specific key performance indicators (KPI’s), as set by the national IAPT team. (22) Essentially all IAPT sites have to submit data to the national IAPT team regarding a number of key areas; access; with numbers of referrals and throughput with numbers entering into treatment, effectiveness and efficiency with numbers that receive treatment within a specific timescale, and numbers of those moving towards employment as well as effectiveness regarding clinical outcomes in terms of attainment of a recovery rate target.

The definition of recovery is through the use of particular psychological measures. All IAPT sites are required to use these. PHQ-9 (23) measuring mood and GAD-7 (24) measuring anxiety are used in every clinical session and the scores at the first and last sessions are used to measure recovery. Patients must score above clinical caseness at first session on at least one measure, and below caseness on both measures at the last session to count as recovered in the IAPT data returns for the KPIs. (caseness = 10 on PHQ9, 8 on GAD7). (22)

2.4 Low intensity psychological therapies

Traditional psychological therapies are delivered face to face with the patient, and usually sessions last 50 minutes to an hour, with a varying range of session frequency and dosage recommendations depending on therapy type, and threshold level of presentation. As mentioned previously, evidence suggests that no one type of therapy is superior. (12). The most improvement in CBT takes place in the earliest sessions, and therefore it can be considered that the whole traditional treatment package has parts that are not necessary and ineffective for clinical gain. Unfortunately not enough is known about which components of therapy are effective for whom, and there is a need for research to focus on the "mechanisms for change". (25)

Scogin et al (16) discuss the benefits and the potential problems with using a stepped care model with depression; however this is in the context of 2003, with only bibliotherapy and pharmacotherapy being the advocated evidenced based low intensity interventions. As discussed previously the range and the evidence of the effectiveness of low intensity interventions has grown since.

Scogin expresses concerns that "severe consequences could occur if severe depressive symptoms (including suicidal ideation) are not promptly addressed with more intensive treatments." At the time of Scogin's publication there was little evidence to support the efficacy of low intensity treatments with more complex presentations, however as mentioned previously the recent meta- analysis found severely depressed patients could benefit from low intensity interventions at least as much as those severely depressed. (26)

It could be argued that in the case of severe depression that impairs daily functioning, low intensity treatment or single strand therapy such as basic behaviour activation delivered frequently may be all that this kind of patient can tolerate and engage with initially rather than a complicated cognitive orientated formulation. Indeed the recommended models of depression for high intensity treatment, including the Beckian Cognitive Behavioural Therapy (CBT) advocate increased behaviour activation initially. Therefore high intensity treatment includes a low intensity style component. Less 'complex' therapies such as single strand treatments e.g. behaviour activation, and low intensity interventions potentially may offer insight

into what are the mechanisms for change, solutions to the unnecessary dosage highlighted by Lovell and Richards (15) and are possibly more cost effective. Behaviour Activation as a stand-alone single strand therapy for depression has been found to be effective. (27,28) Low intensity interventions such as internet or telephone based CBT, and guided self-help are psycho- educative in nature, brief, with usually between four to eight sessions, and have demonstrated their effectiveness. For instance, Hammond et al (29) found low intensity telephone CBT based intervention to not be inferior compared to face to face CBT, except with those with more severe illness, where traditional CBT was superior. However a meta-analysis of 16 studies looked at the effectiveness of low intensity interventions on more severe depression. It found that more severely depressed patients can gain at "least as much clinical benefit" as those with less severe presentations. (26) Further research (13) support this, and suggests there is a cost benefit, "These treatments have been found to be effective in the treatment of depression, with comparable effect sizes as face-to-face therapies, while needing fewer resources". Certainly the evidence makes low intensity treatments attractive economically and clinically, achieving the same outcome with less resource, and within a context of a stepped care model, provides support regarding the principle of least intrusive intervention first. The national IAPT programme provides a workforce and clinical delivery framework which advocates the use of low intensity interventions, and recommends a stepped care approach in line with the NICE guidelines. (19)

2.5 A brief overview of delivery models of stepped care

Mental health psychological services treatment delivery in the UK is generally informed by NICE guidelines. As discussed previously the current NICE guidelines offer a mixed stratified matched and pure stepped care model, for depression advocating least intensive treatment first, with some variation for some anxiety disorders, and some exceptions e.g. PTSD where current recommendations suggest step 3 interventions as a first line treatment. (figure 1). Bower and Gilbody (18) neatly describe the definition of stepped care, clarifying that the principle of 'least restrictive treatment' is often defined as the minimum dosage and treatment intensity required for the most optimal outcome, Bower et al (26) further suggest that "it is legitimate to

include low intensity interventions in the first step of a stepped care system and to encourage most patients to use them as the initial treatment option, even when the initial severity of depression is high."

As discussed previously, the development, definition and inclusion of stepped care within NICE guidelines occurred prior to the development of IAPT. Simultaneously there appears to be an expansion of the definition, with a number of slightly different models of delivery, all being described as stepped care.

There appears to be a version of stepped care, which has different layers of intensity, however patients are allocated directly to a step, for a treatment where the intensity of treatment matches the intensity of complexity and severity. This is a stratified matched or allocated model of stepped care. A second version is where the least intrusive intervention principle is applied more stringently, and patients are treated at the lowest step first, then 'stepped up' if needing more intensive treatment. This is true, or pure stepped care.

Some literature also includes the collaborative care model as a version of stepped care. Collaborative care is usually defined as multi component, with a number of professionals around the patient, e.g. "a medical doctor, a case manager (with training in depression and anxiety), and a mental health specialist such as a Psychiatrist". (30). This kind of arrangement would usually be seen in secondary care, over a longer period of time than primary care delivered brief therapy. Given that "a key aspect of collaborative care is 'case management' (Gilbody 2003a)" (30), this places collaborative care usually in the threshold level of secondary care mental health provision. Some models of collaborative care are delivered with a variety of threshold levels for mental health and comorbid with other problems, for instance and long term physical conditions (LTC) such as Asthma, or Diabetes. A meta-analysis found collaborative care to be an affective model of care with LTC and depression. (31) "The collaborative care model is based on the principles of chronic disease management applied to conditions such as Diabetes." (30) Here the physical problem is chronic and longstanding; however the mental health problem may only require brief therapy. It is primarily targeted to patients with specific physical care management needs; therefore the focus of the model delivery is this, with the mental health treatment as a component, rather than the main focus. Collaborative care may

contain the points of treatment review and adjustment, and a component of step up or sideways, fitting with the broad definition of the stepped care model, however as delivery is not necessarily comparable in terms of focused problem, threshold level of mental health problem and treatments offered, there is a counter argument regarding validity of comparison with other stepped care models.

Excluding the literature focusing on collaborative care, models of treatment delivery appear broadly differentiated as allocated /stratified or pure stepped care (18, 32) Whilst there has been some attempt within the literature to define more clearly what stepped care service delivery actually means, It would appear that both in policy and routine practice there are different interpretations being delivered and the evidence of the effectiveness of these various interpretations is limited.

One RCT study in the Netherlands aimed to compare different service models of matched care and stepped care, defining the latter "in a stepped care model all patients start with an evidence-based treatment of low intensity as a first step". (33) Interestingly its first step of treatment is either brief therapy, which broadly appears to compare with the guided self-help low intensity intervention in England, and CBT sessions equivalent to high intensity step 3 treatment in the U.K, particularly within an IAPT service. This Randomised Controlled Trial (RCT) set in routine practice, and although found in favour both clinically and cost effectively for the stepped care model, unfortunately the study did not find statistical significance, suffered with a large number of drop-outs, and lacked power.

An observational study (32) (which is critiqued within the literature review, section 4) attempts to analyse different models in routine practice in 4 different services. Unfortunately large levels of missing data and lack of availability of clinical outcome data meant that it was difficult to reach any conclusion. However this study highlights that the "variation in models was significant", and that service design appears to be responsive to resource capacity, different local definition of need or interpretation of the NICE guidelines. The study (34) suggests that whilst a combination of "stepping and stratification is likely to be required, the relative importance of the two different mechanisms was not explicit in the NICE guidelines." It further states there is limited literature pertaining to the evaluation of effectiveness of the stepped care approach, and "significant questions remain about implementation", although goes on to define

the difference between ‘stepped’ and ‘stratified’ in line with Bowers and Gilbody (18), it states that it is not explicit in the NICE guidelines the important difference between the two. This is the fundamental problem with interpretation of the guidelines and implementation in routine practice.

Furthermore, most literature is prior or leading up to the development of IAPT, therefore are more general to mental health teams, psychology services and collaborative care. Most are consistent with the idea that stepped care values the idea of low intensity and least intrusive intervention, and the stepped care model is ‘self-correcting’ in that where patients fail to progress they are stepped up. (18, 32) However what is also describes is the notion of targeting the ‘right’ treatment that achieves the most benefit. (32)

Whilst in theory, combining the two principles may be desirable, they could be contradictory, as still not enough is known about what components of treatment are most beneficial and works with whom in stepped care. (32) The idea of choosing the right treatment is limited, and surely if enough was known about what constitutes the right treatment for each patient, the mechanism of self-correction with the stepped care model, i.e. stepping up with failure to progress, would not be necessary, and simply the stepped care model of implementation would be a stratified one.

Returning to point about not enough is known about what works with whom, (32) it could be reasonable to assume that there are numbers of patients that present as severe and complex that are allocated and may receive more dosage of treatment than they really need, simply because they are allocated that level of stepped care treatment. This was indicated within the MAPLE paper (15) as an argument for an increase of use for low intensity treatments and a pure model of stepped care.

If not enough is known about what works for whom then there is a question about how allocation of treatment is decided. If therapists follow the NICE guidelines and specific disorder model treatment plan of recommended number of treatment sessions, this will average 12-16, depending on disorder treated regardless of individual patient dosage need. Nice guidelines describe the allocation of increased intensity and dosage of treatment delivered by more specialist staff, matched to the increased severity, complexity, chronicity and risk of patient presentation. (10)

Within IAPT, the psychological measures used that measure frequency of symptoms are self-reporting, therefore open to some level of personal interpretation and therefore variation. Alongside an assessment, the clinician is likely to use this information to inform the decision regarding complexity and severity. Again this is subject to personal interpretation and variation, within services, and across the country.

Other elements that may inform decision to step up is dosage of treatment, and more subtle and less tangible factors regarding therapist, i.e. inexperience or lack of skill or knowledge with particular disorders, interpersonal processes, i.e. the communication, connection and relationship between patient and therapist, and generally therapist bias. This notion suggests a possibility in a system of matched care open to the influence of therapist decision, personal preference of the therapist may be one of the driving factors to allocation of intensity of treatment, rather than specifically what may be needed.

NICE guidelines are developed through a consideration of evidence, and in practice the equation of complex and severe needs more complex and intensive therapy would appear logical. This is what is recommended within the NICE stepped care guidelines. However this may cause a pragmatic and resource problem, if a service receives the majority of their referrals that fall into the more severe, complex, chronic and risky presentations, then by this equation more patients would be allocated to higher intensity interventions. Given the findings of the Layard's (3) Department of Health (DOH) report it would appear that this was indeed the case that led to long waiting times for patients accessing a talking therapy. A wish to tackle this problem is the principle of the national IAPT programme.

More specific to IAPT, a review of the national IAPT first year implementation and explore differences in stepped care treatment levels in terms of which patients are most likely to receive a high intensity treatment. (21) Findings were that higher PHQ9 or phobia scores are slightly more likely, with higher GAD 7 and women more likely. Older patients are more likely to have some treatment but less likely high intensity or CBT. Clark (14) reviews the progress of the national IAPT programme in its third year at the time of publication. It makes reference to a Department of Health (DOH) report that states the majority of patients are receiving NICE compliant treatment, if the

psychological measures score profiles of other IAPT services are similar to this study sites, then one could assume that the majority of services are using an allocated and stratified model. However a prospective study of the IAPT pilot sites by Richards & Borglin (34) discusses that most patients are treated through low intensity interventions, and a smaller number are stepped up to high intensity. Therefore it does not appear clear whether there is consistency amongst the IAPT sites nationally regarding a definition and implementation of a stepped care model. Neither does there appear to be consistency in the literature regarding the decision mechanisms for stepping up, whether it be clinician judgement, psychological measures score thresholds, types of disorder, failure to reach agreed goals, or patient choice. (34)

There appears to be an assumption that the basic definition outlined by Bowers et al (18) and Richards et al (32), is implemented according to both the literature, NICE guidelines and in routine practice. There are a number of questions and gaps of evidence within the literature: not enough is known about which treatments work best for whom, and how many treatment sessions should be delivered. Which then does question the efficacy of a matched care model, and perhaps this is more directed by therapist preference and the notion that complexity and severity should always be treated with high intensity therapy. Not enough is known about the optimal implementation model of stepped care, the number of steps and the best range of treatment, and optimal step up choice points. (32)

These pose problems for routine practice, where services are responsible and accountable for the safe care and treatment of often vulnerable people, and are charged with the task of improving patient's mental health difficulties and wellbeing. For services such as IAPT's delivering psychological therapy to achieve those aims, and publicly measured on their performance of such, this is a particular area of interest to know what works and how it needs to be delivered.

There are many factors involved in potential analysis of what works for whom, with a few mentioned briefly above however this is beyond the scope of this study. In the absence of comprehensive evidence about the efficacy of components of therapy, mechanisms of change, intensity and dosage of treatment, that can inform and direct decisions without bias regarding allocation of treatment to a specific patient, if a

service delivery model can improve outcomes then this is worth considering until such time bespoke treatments are able to be developed.

This study site is a large IAPT service within the North East of England, and through its move from one type of delivery model to another has provided an opportunity to explore the impact of different stepped care models of delivery.

2.6 Research study rationale

Although research demonstrates the effectiveness of psychological therapies with common mental health problems, as briefly mentioned earlier, there is no superiority between modalities of therapy. (12, 13) Furthermore we do not understand what components of therapy are the most effective and whether this varies by the patient presentation. There is perhaps an assumption pervading routine practice that the more complex and severe the presentation of a patient, the more intensive and complex the therapy, with a larger number of sessions is needed, when we do not understand optimal dose or intensity levels for different patients. There is some evidence within the literature that low intensity interventions can be as effective as high intensity regardless of complexity and severity (27) and are potentially more cost effective. There appears to be different interpretations of how to define and implement a stepped care model, and this may dictate the emphasis on throughput to low or high intensity interventions, yet there is not enough known about how particular interpretations and implementations of stepped care and therefore service design can impact on clinical outcomes. Specifically there is a gap in evidence regarding the efficacy of an allocated model compared to a progression model of stepped care, and a possible continued belief perpetuated in routine practice individually and systemically that severe presentations should only be treated with a high intensity therapy. Through a change in model delivery, the study site as described in more detail in section 5.5 provides a unique opportunity to explore some of these issues.

This study firstly set out to explore the current literature regarding the evidence for the efficacy of stepped care, and secondly through an observational cohort design analysed the clinical outcomes with two variations of stepped care as delivered over a 4 year period by an IAPT service. By also isolating those scoring moderate to severe allows exploration of association of severity, outcome and model, to explore

the anecdotal notion that those with severe presentations would have poorer outcomes if treatment was not matched.

Chapter 3: Literature review search strategy

An initial exploration of the literature was undertaken to scope the evidence of the key issues surrounding the research question, and indeed if there were any studies that specifically relate to the question of comparing patient outcomes with different models of stepped care delivery. This initial exploration informed and refined the study aim, to explore what impact, if any change in service delivery design has on clinical outcome.

In order to benchmark any results of a particular study, a review of relating RCT's, any other studies, and any other relevant literature is undertaken. In order to fully appraise the literature relating to system design, a full systematic review and meta-analysis where possible would be conducted, however to do so falls outside of the practical limitations of this project. In order to retain the rigour of such approaches key elements of a review have been incorporated into this study's literature review. Systematic searches were conducted across four databases most likely to include research relating to the study question.

The Cochrane handbook for systematic reviews (35) describes the purpose and method of a systematic review to answer a research question, with the collation of "all empirical evidence that fits pre-specified eligibility criteria", in an attempt to minimize bias uses clear methods. Systematic reviews reduce the risk of bias in research findings by following a transparent approach to study identification, assessment and data extraction. Following the PICO approach, population, intervention, comparison, outcomes sets out this out in a methodological clear manner. (37) It is also important is to decide whether only published studies will be included and what kind, e.g. RCTs, qualitative research, and whether there will be studies in any language accepted or if restrictions will apply.

3.1 Population

In order to minimise any selection bias the criteria should be wide enough to include enough studies, but specific enough to ensure validity and relevance within the scope of the research. For instance the research area concerns an adult population, and generally adult is defined as over 18 years of age. However in terms of mental

health service criteria there is a grey area with regards to the age definition and type of service. For instance there would appear a general custom and practice for Child and Adolescent Mental Health Services (CAMHS) to deliver services to those up to aged 16, or 18 where they are in full time education or in the looked after system with the Local Authority. Early Intervention with Psychosis teams generally see patients from aged 14 – 35. How stepped care is delivered if at all may differ between children's and adults services, and also the nature of interventions within children's service model delivery may be different to adults therefore providing too many variables for the scope of this study. The range of disorders are usually separated by severity for service delivery organisation, e.g. primary care for common mental health problems, secondary care for more severe, perhaps at risk of harm to self or others or needing care co-ordination (a package of other services involved co-ordinated by a lead professional) and tertiary services offering more long term involvement, or short or long term hospital care. The range of literature that covers all levels may be broad and variable. Also specific disorders may have differing levels of intervention and therefore not directly comparable. IAPT services are offered to people aged 16+ and there is no upper age limit. In terms of the validity of the results of this study it is important for the population group within the literature e to be similar. Therefore the population to be targeted in this review is adults with common mental health problems (anxiety or depression) treated within psychological therapies services.

3.2 Intervention and Comparison

"The second key component of a well-formulated question is to specify the interventions of interest and the interventions against which these will be compared (comparisons). In particular, are the interventions to be compared with an inactive control intervention (e.g. placebo, no treatment, standard care, or a waiting list control), or with an active control intervention (e.g. a different variant of the same intervention, a different drug, a different kind of therapy)?" (35)

The literature pertaining to the effectiveness of psychological therapies is well evidenced and the intervention under consideration here is not a specific type of therapy, but a model of delivery of psychological therapies.

The main model of delivery is stepped care, defined by Bowers et al (26), NICE (10) and Richards et al (32) however within the literature there are a variety of models of care provision that may have a stepped care component, be a component of stepped care, another term of description, or a different model of delivery for comparison. Therefore initial searches covered all known terms and phrases that are relevant to the question, for instance 'stepped care', 'allocated care', 'matched care', 'collaborative care', and 'stratified care'.

3.3 Outcomes

RCTs remain the gold standard with the regards to reliability of evidence, In RCT's treatment protocols are followed to ensure that the core components of treatment are the same to ensure standardisation and validity of outcome. Confounding variables are acknowledged and if possible controlled, in an attempt to ensure the results are reliable. However although widely acknowledged that RCT's are of the highest standard with regards to accepted reliability of evidence this method of study is not without its flaws. (36) RCT's are a deductive method, which supports the notion of gold standard, as demonstrable causality is better than degrees of association. Generally what the outcomes of RCTs demonstrate is that the results are applicable to the population criteria of the RCT, there is potentially an issue with external validity, unless it is demonstrated that the study sample is indeed representative of the general population. However the very nature of the RCT, controlling variables will mean that the study sample is likely to be narrower in clinical presentation compared to a routine practice population. They can be challenging to deliver, in terms of recruitment of participant numbers to ensure the study has enough power for valid results, and It may not be replicable in routine practice, through specific variables not able to be isolated, for example patients with co-morbidity. There are other methods to evaluate and examine data within routine clinical practice, one of the most common criticisms of observational studies is that they risk bias with the over estimation of treatment effects and therefore are less valid, however one review found that "the results of well-designed observational studies (with either a cohort or a case control design) do not systematically over-estimate the magnitude of the effects of treatment as compared with those in a randomised controlled trial on the same topic."(37) The benefits of observational studies are that they are more applicable to routine practice, as it demonstrates outcomes within settings with a

number of variables, and therefore could be viewed as more realistic. Given the context of this study taking place in routine practice it is absolutely valid to review studies whose methods may be more comparable, other than RCT's.

The Cochrane Handbook states "Review authors should consider how outcomes may be measured, both in terms of the type of scale likely to be used and the timing of measurement. Outcomes may be measured objectively (e.g. blood pressure, number of strokes) or subjectively as rated by a clinician, patient, or carer (e.g. disability scales). It may be important to specify whether measurement scales have been published or validated. When defining the timing of outcome measurement, authors may consider whether all time frames or only selected time-points will be included in the review." (35)

In the mental health field, outcome of intervention delivery may be multi-faceted, with observable, measurable aspects which may be open to interpretation, bias and confounding variables. For instance clinician's observations and patient verbally reporting an improvement in symptoms are valid to those individuals, however subjective in terms of whether that demonstrates an intervention to be effective and generalizable to a wider population.

The outcome of interest for this study is the effectiveness or the results of implementation of a particular service delivery design. Therefore literature of interest will be in the first instance systematic reviews and published RCT's regarding the effectiveness of stepped care delivery of psychological therapy with anxiety and depression. "By providing a reliable synthesis of the available evidence on a given topic, systematic reviews adhere to the principle that science is cumulative and facilitate decisions considering all the evidence on the effect of an intervention". (37) Secondly given that the method of this study is observational, any observational studies regarding the efficacy of stepped care were also included.

3.4 Inclusion and exclusion criteria

Using the PICO approach, studies included were those published in English, systematic reviews and meta-analyses, randomised controlled trials, and observational studies in routine practice. Participants were adults with mental health problems treated in a psychological therapy service, within a stepped care service

design. Models of service design were searched for using the following ‘stepped care’, ‘allocated care’, ‘matched care’, ‘collaborative care’, and ‘stratified care’. The focus of studies included had to be regarding the effectiveness of the stepped care model.

The Cochrane library search of systematic reviews

anxiety or anxious "stepped care" or "collaborative care" or "stratified care" or "matched care" by record title.

Table 1: Initial search inclusion criteria using a PICO approach

Search terms	Filters	Limits	Results	Manual Filter	Results	Duplicates	Unique results
(mental health) OR anx* OR depress* AND therap* AND (stepped care) AND effect* OR impl*	Title, abstract, keywords	1994-2014, systematic reviews, Depression, anxiety and neurosis group	119	By title and abstract using inclusion / exclusion criteria	3		3
(mental health) OR anx* OR depress* AND therap* AND (matched care) AND effect* OR impl*			119		3	3	0
(mental health) OR anx* OR depress* AND therap* AND (allocated care) AND effect* OR impl*			119		3	3	0
(mental health) OR anx* OR depress* AND therap* AND (stratified care) AND effect* OR impl*			119		3	3	0
(mental health) OR anx* OR depress* AND therap* AND (collaborative care) AND effect* OR impl*			130		3	3	0

As well as systematic reviews searched for through the Cochrane database, several other established health databases, psycINFO, MEDLINE, CINHAL, and EMBASE were also searched using similar key terms, adjusting where necessary according to the specific coding nature of that database. (Table 1).

The searches altogether returned 152 papers. Duplicates were removed, and the manual filter was applied. Through my professional qualification and experience working within the field I am aware that the research in the area of the effectiveness of psychotherapy with anxiety and depression is broad and extensive, and particularly rich in terms of evidence pertaining to Cognitive Behavioural Therapy. The study question does not relate to contrasting and comparing different types of therapies for mental health problems, and is beyond the limitations of this study to do justice to this topic. The question relates to the effectiveness of a particular service delivery model.

There is a variety of definition and implementation of service delivery model in routine practice, the focus of this project is the exploration of definition of “stepped care” and its evidence of effective implementation in routine practice. Initial searches included collaborative care, and resulted in a number of trials, and one Cochrane systematic review (32) that focus specifically on the delivery of collaborative care. The literature regarding collaborative care was analysed briefly to ascertain its relevance to this study. There are a number of studies that focus on the delivery of collaborative care based on a psychotherapy model and Archer et al (30) demonstrate the effectiveness of collaborative care in improving Depression. One of the difficulties is that the collaborative care is often delivered within or around a stepped care model, therefore is it quite difficult in this case to identify either as a separate model of delivery. Whilst this is a similar case for stratified or matched care models i.e. services are likely to be describing their model of delivery as stepped care but more likely to be delivering a combination of that and matched or stratified. As described previously to be able to differentiate between studies and identify those that are based on a psychotherapy stepped care model, and delivered within a threshold level of primary care is a systematic review in itself.

Collaborative care is certainly a model that is well evidenced regarding depression, and there is certainly a need to increase and review the evidence for collaborative

care with LTC and mental health problems, and anxiety (30) but this is beyond the scope and specific focus of this study.

Given its main focus is either at a secondary care threshold level or the physical problem is the identifier for this model of care, this legitimises collaborative care to be excluded from the main literature review for this study.

Removing any results that covered collaborative care, adding together papers already obtained by the author prior to the commencement of this study, and subsequent papers published after the time of the initial database searches, brought the total meeting the criteria of the search question to 8 papers.

Two of those are systematic reviews (38, 39) and between them included within the review a number of studies identified as relevant to this research question and included in the 8. (20, 33, 34, 40) Therefore to avoid duplication this study will not undertake an analysis of those papers but concentrate on the analysis of them within the systemic reviews. Richards & Borglin (34) was not included in the van Straten (39) review as it did not fit the criteria of being an RCT. The Richards et al (32) paper did not feature within the Firth (38) review, and it is not clear why given it is an observational study exploring the effectiveness of stepped care in routine practice, and the authors' inclusion of other observational studies. The paper is highly relevant to this study; therefore it is included in the literature review. This then brings the total number of papers for the context of the specific question around effectiveness of stepped care service delivery, to 4, to be reviewed in the next chapter.

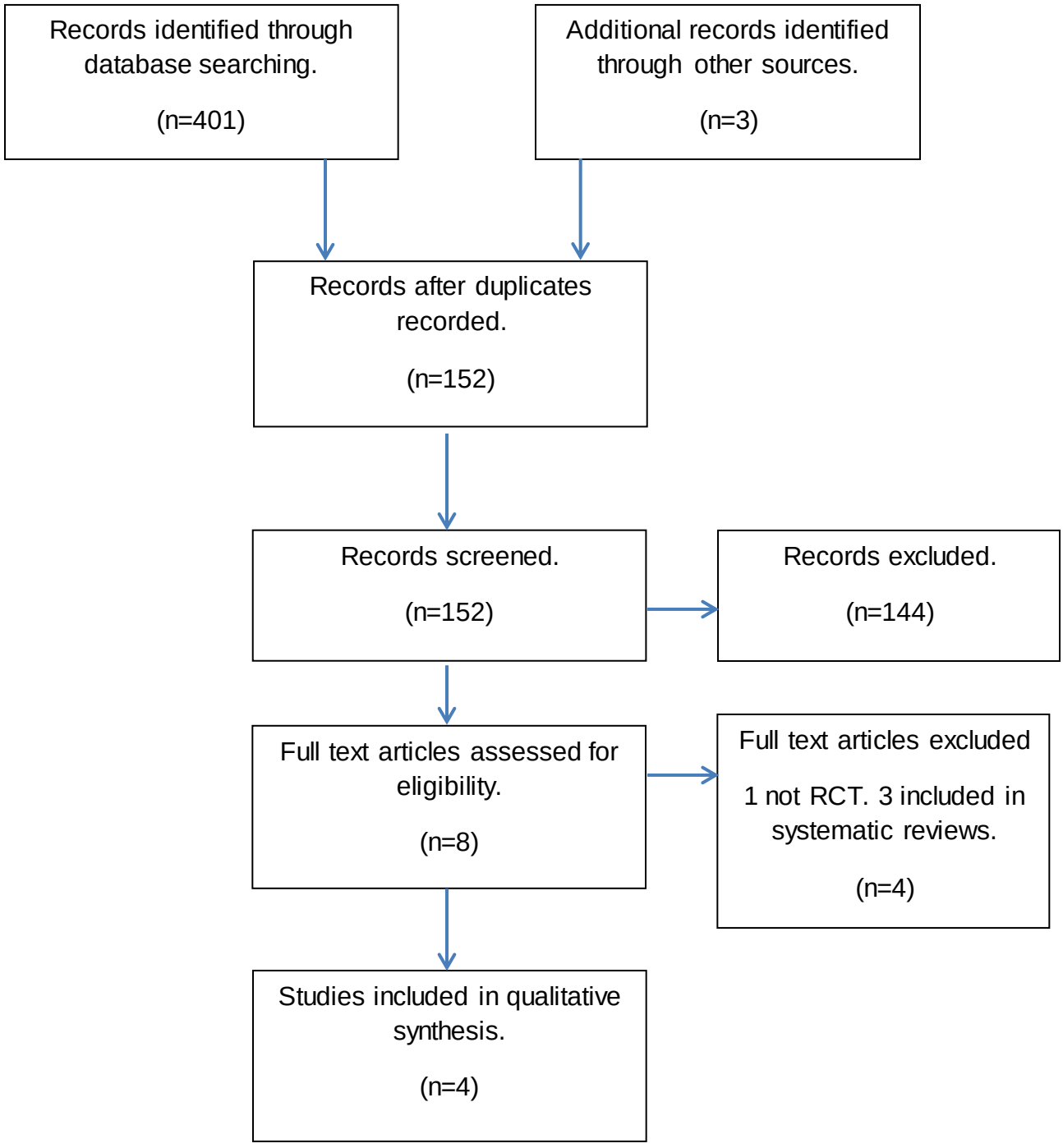


Figure 2: A search strategy diagram

Chapter 4: Literature review findings

The studies were reviewed using either the CASP cohort study checklist (41) or the PRISMA checklist for systematic reviews. (42)

4.1 Delivering stepped care: an analysis of implementation in routine practice

David A. Richards, Peter Bower , Christina Pagel, Alice Weaver, Martin Utley, John Cape, Steve Pilling, Karina Lovell, Simon Gilbody, Judy Leibowitz, Lilian Owens, Roger Paxton, Sue Hennessy, Angela Simpson , Steve Gallivan, David Tomson and Christos Vasilakis. (2012)(32)

Findings

Part of a larger project aiming to "develop a decision and modelling aid for services designing stepped care organisational systems." , this is an observational study using data of patients in 4 different routine health settings operating stepped care models, with an analysis of the proportion of patients accessing treatment, at which 'step' and the transitions between steps.

The study found that the interpretation of the NICE guidelines for stepped care was implemented with large variation across all 4 sites, with a particular difference with the ratio of low or high intensity treatments received by differing proportions of patients across all the sites.

Analysis

The study positively met the criteria outlined in CASP for all 12 questions apart from question 6 regarding follow up of participants, which was not relevant in this instance. Consideration of quality, validity and bias of results and methodology as outlined in CASP are further outlined.

The context and rationale for evidence based practice for psychological therapies with anxiety and depression is briefly acknowledged, and the authors highlight that the stepped care model within NICE guidelines was not based on the same rigorous evidence synthesis. Therefore this study poses the question regarding the evidence for stepped care aiming to answer the following questions:

"1. What different models of stepped care are implemented in practice?

1. How do patients access and flow through the different models?
2. What proportion of patients are managed at each step, and what proportion 'stepped up'?"

The authors state that there was no analysis of outcomes, and acknowledge this is a limitation to the findings, however identify that outcome data across all sites was extremely limited, and not routinely collected by information systems or clinicians, which as the authors acknowledge is a great concern. Although the use of psychological measures is not the only way to assess presentation and progress, they can aid clinical judgement and decision making, and if tested as highly reliable and valid can give some objective perspective which is easier to quantify for research purposes.

In the setup of this study, the authors were involved in a 'consensus workshop' in which they provided facilitation to the services to build their version of a stepped care approach. The authors state there was no influence regarding new service design, however the methodology of this is not outlined in detail, nor any mention of a qualitative study method which would discuss the concept of impartiality, or influence and whether there is such an occurrence through any interaction from the researchers regarding the service model design.

Essentially the data regarding the patient care pathway flow is analysed, and the 4 sites compared. The method of data analysis was using categorical data and continuous data as means, and standard deviations. (32) There were a large number of patient cases excluded from the results due to missing data, however throughput data of nearly 8000 cases was reported on. The authors acknowledge that the time limitations of the project meant that there was no endpoint data, which is disappointing as there is a need to evidence the clinical effectiveness of a stepped care process. The population characteristics analysis shows one particular site ranking high on the Index of Multiple Deprivation, as there is no analysis of the clinical information it is not known whether the level of mental health problems differ from the other sites, which is possible given the evidence link between poverty and mental ill health. (43)

All four sites had different referral and assessment management systems, with different types of professionals assessing, all which potentially effects patient flow, and clinical decision making. Essentially the models of delivery were either stratified with a matched care/allocated system or stepped care. One site deployed a much larger amount of resource of experienced staff to assessment to support allocation to the 'right' treatment level, however this resulted in the service not being able to meet the demands at a high intensity level and therefore patients were shifted to low intensity interventions. It is disappointing that there are no clinical outcomes to compare all the sites, however in particular it would have been interesting to see the outcomes of those assessed as in need of high intensity, but treated with low intensity. It is indicated in this study that a stratified (allocated) model of delivery appears more likely to have a large volume of patients allocated to high intensity treatment, and a stepped care model will likely have the inverse. The authors acknowledge the limitations of the study, particularly with regards to missing data and inability to analyse clinical outcomes. There is a comment on the possibility of clinician bias, with "triage or assessment by a professionally qualified workforce may lead to more people receiving high intensity", which then may then also be reinforced by service design and policy. The issue of heterogeneity of interpretation and implementation of stepped care is discussed and the authors highlight that the NICE guidelines do not provide a "formal blueprint for the organisation and delivery of services" which may contribute to the variability of stepped care definition and practice implementation.

4.2 Service use, drop-out rate and clinical outcomes: A comparison between High and Low intensity treatments in an IAPT service

Stella W.Y.Chan & Malcolm Adams (2014) (44)

Findings

Using routine care data from an IAPT service, this is a small sample study comparing low and high intensity treatments, analysing outcomes and drop out rates.

Using a between groups design, it takes a small sample of 100 cases from an original dataset of 15,082.

It finds that there is no difference between groups regarding dropout rate, a level of % difference between groups in terms of recovery rates – 50% high intensity, 55.3% low intensity, however no significant statistical difference.

Analysis

The study positively met the criteria outlined in CASP for all 12 questions apart from question 6 regarding follow up of participants, which was not relevant in this instance. Consideration of quality, validity and bias of results and methodology as outlined in CASP are further outlined.

The study sets the scene of the national picture in terms of recovery rates averaging at 42% as reported by the "latest national evaluation study". This the authors are referring to an evaluation of the first year of roll out published by NEPHO, 2010, and although the authors also acknowledge other studies that present variable recovery rates, it would have been useful to have comment on the potential variable that may affect recovery rates, such as different time frames of studies with different groups of staff that in later periods may have more experience. This study sample is drawn from a dataset covering a time period of nearly 3 years, which includes the first year of IAPT roll out where potentially a larger number of therapists were in training, and then in subsequent years may be more experienced. With 100 cases samples from a 3 year timeframe, it would have been helpful if there was an explanation of the level of competency or experience of therapists delivering which may have an impact on outcomes.

The study begins with a large dataset, 15,082 cases, which would be a healthy large number to analyse. Firstly the authors state that each client contact was recorded as a separate entry, which is true for most health databases, however they state this then gave 88,072 entries to analyse, which they state is too large and this is the rationale for a much smaller sample of 100 cases. It is difficult to understand this as data extraction regarding outcome and drop out could have been extracted through a cohort of cases which could have been on a case basis rather than individual clinical contact entry. This cohort could be smaller than the whole dataset, but larger than 100 to ensure generalizability. The authors state that they arrived at the sample size using a general statistics principle; however "a formal power calculation was impossible due to the unavailability of relevant data in the existing literature." It is unclear as to what the authors are meaning here and an assumption can only be

made that they are referring to the fact that IAPT is relatively new, there isn't enough studies, particularly randomised controlled trials with which to base a prediction of effect size. However there is an abundance of studies that explore the efficacy of CBT, and more lately low intensity interventions, both of which occur in this study therefore this information could have informed the consideration of a formal power calculation.

Secondly the authors state that the intensity level of which cases received treatment was not explicit in the original dataset, and in order to remedy that they separated cases out into low and high intensity according to the pay band of the therapist delivering treatment. This is a reasonable step to take, however again it indicates potentially issues in the services clinical recording system design regarding input of information that can be collated. Both the issue of service system design and therefore data input accuracy and validity issue and the lack of formal power calculation weaken the results.

The authors also go onto acknowledge limitations with the randomisation procedure where they did not control the higher probability of certain cases more likely to be selected through their greater number of contacts. Whilst more sessions of therapy does not necessarily mean a better outcome, the fact that this variable is not controlled means that the higher recovery rate may have been influenced by the number of treatment sessions.

There are other variables that would have been useful to have explored. The authors explain the context of the service with the standard description of the IAPT criteria as generally those with mild to moderate symptoms of anxiety or depression, and operate a stepped care model, describing the usual definition of low intensity to the majority, with step up for those not making progress, and also the exception of those clinically judged to be more severe being offered a higher step in the first instance. The authors rationalise their purpose for this study from the findings in Gyani, Shafran, Layard and Clark (11) where it is suggested that "services that made better use of stepped care produced better outcomes, and further estimated that an improvement on the "step- up" rate could potentially increase the recovery rate to as high as 54%", further stating that as there are large variability amongst the IAPT services reviewed here, it is important to research the differences between high and

low intensity treatments within the same service. This is worthy of exploration, which the study does, however the step-up issue is not explored fully. There is no explanation of the ratio of those receiving low intensity to those receiving high intensity in the first instance, so there is no way of knowing if indeed the majority of referrals to this service do in fact receive a low intensity intervention in the first instance, or whether this may be the broad description and expectation of the service however in reality severity of presentation and clinician interpretation may impact on the actual proportion receiving low or high interventions. As this is not discussed, appear to be considered or controlled in the methodology, this may also have an impact on the types of cases used study sample, and therefore may affect the results. As previously mentioned, the intensity of treatment received was not documented, nor was the ratio of step up from low to high, where longer dose of treatment is received and the presentation likely to be complex, and less likely to recover. The authors do discuss how this may have 'dragged' down the recovery rate in high intensity, stating that lack of 'stepping up' data meant an inability to test this possibility. The authors describe in their methodology how they use pay bands of clinician treating to differentiate cases into low and high intensity, therefore it should have been possible to undertake an analysis of sample groups clinical recording entries which could have given information about which type of clinician was treating each patient at the commencement and end of treatment, and where it differed this should be able to identify those cases that have stepped up.

The authors describe a standard definition in IAPT of stepped care to indicate how cases are 'allocated' to each step, however there is no analysis of score by step, therefore it is not clear whether score severity correlates with this definition.

What is really interesting, in light of the above is there is no difference in the two groups, low or high intensity with baseline scores for all cases. Therefore the reasons for allocation in the first instance to a high intensity treatment must be dependent on something other than scores, which are self-reporting and therefore an element of subjectivity, however other information and clinician judgement are the deciding mechanisms in the allocation of intensity of treatment, however these factors are not controlled in this study and therefore it is unknown whether those cases receiving low intensity treatment differed with complexity or co-morbidity which may impact on outcome of treatment.

An important summary of this study is that it found that high intensity deliver more treatment sessions, but the recovery rate is lower than those treated at low intensity, and yet there was no difference between the groups in the baseline psychological measures, questioning whether score severity can be an indicator of step intensity needed. The authors acknowledge this surprising finding and that it is possible that both groups could contain similar presentations, which could imply that "high intensity does not offer additional benefits." Conversely it is also possible that although no difference in scores, clinical presentation was different and the high intensity group more complex and chronic. Further analysis of such would be useful.

In summary, the results could be generalizable to the clinical population of IAPT and show some consistency with efficacy of step 2 interventions as demonstrated in other research. (26) However there are indications of possible issues with the service data collection design and the study methodology- some of which is acknowledged by the authors, which may question the reliability of the results.

4.3 Stepped care treatment delivery for depression: a systematic review and meta-analysis

A.van Straten, J. Hill, D.A. Richards, and P. Cuijpers. 2014 (39)

Findings

This is a robust review of 14 studies, (n=5194, 2560 in stepped care) of which 10 are used in a meta –analysis (n =4580, with n= 2243 in stepped care).

7 studies are regarding the delivery of a collaborative care model, 6 studies are regarding an increasing intensity of stepped care with just 2 studies with progressive intensity of stepped care.

The review finds that stepped care has a moderate effect on depression with the progressive intensity doing significantly worse.

Analysis

This systematic review positively met all the criteria of the PRISMA checklist apart from there was no information regarding any protocol and registration or funding.

Whilst the methodology of this review is of high quality, the definition of stepped care for this review and therefore the inclusion criteria is by the authors' acknowledgement arguably too wide and so it does raise issues regarding validity of comparison of studies with extreme heterogeneity of treatment delivery organisation.

However there is a dearth of studies regarding stepped care, and given the demonstrable difference with interpretation and implementation of stepped care in the previous study (32) of which one author is also involved in both studies, it is also an understandable point to include other treatment delivery systems such as collaborative care to compare a wider range of delivery systems.

The validity was assessed using the Cochrane handbook criteria, and 2 reviewers undertook a quality assessment of each study independent of each other most studies were rated of good quality. Publication bias was also tested, and none found. The meta-analysis included between-group effect sizes, and sensitivity analyses, where a study reported more than one outcome, effect sizes were also pooled, and heterogeneity tested.

The authors define the inclusion criteria as quite wide ranging, covering adults, with a diagnosis or symptoms scoring above a threshold for depression, 'stepped care' as one of the trial groups, includes a psychological therapy, availability of more than one treatment that 'step up' based on 'systematic clinical evaluation' at a specific time point. It also included studies with physical and psychiatric co-morbidity. This wide definition resulted in the inclusion of a number of IMPACT American trials, IMPACT being a model of delivery akin to collaborative care. The definition also included pharmacological treatment, and 'step up' was loosely defined as an adjustment in treatment. There was heterogeneity between studies regarding number of steps, types of treatments offered and length of intervention.

The 7 collaborative care studies and one other, involving psychological treatment and anti-depressant medication had no progression of increasing therapeutic intensity, rather there were review points and treatment adjustment. Whilst the authors found that stepped care had a moderate effect on depression, the 2 progressive stepped care studies demonstrated a worse effect than those without a clear intensity order. However the authors do acknowledge that with only 2 studies the results need to be treated cautiously, and perhaps consideration needs to be given to the fact that one of these studies was underpowered, (40) and the other relied in two levels on anti-depressants as well as a psychological talking therapy and results may be more reflective on the variability of effect of pharmacotherapy over

the long term if discontinued or not adhered to. (45) The study found that the stepped care 6- month effect size was similar to collaborative care as found in the Cochrane review. (30)

The authors do acknowledge that their definition of stepped care is “debatable”. There is already a systemic review on collaborative care, (30) and therefore arguably no need for another. Although a more strict inclusion criterion would have resulted in a smaller systematic review, even this would have demonstrated a wide range of heterogeneity. As it is with the inclusion of collaborative care and also studies that include physical conditions, the heterogeneity is extreme, and arguably either weakens the findings, or demonstrates the perspective that stepped care definition is open to wide interpretation. They do suggest that more studies of ‘true’ stepped care need to be undertaken.

The authors also reviewed the cost analyses where available, and found that there is an indication that stepped care interventions might be more cost effective, however there are still huge gaps in the evidence, recommending that stepped care and matched care, or high intensity only need comparison and cost effectiveness measured.

4.4 The clinical effectiveness of stepped care systems for depression in working ages adults: A systematic review

Nick Firth, Michael Barkham, Stephen Kellett. 2015. (38)

Findings

This systematic review analyses a total of 14 studies to evaluate the evidence of the effectiveness of a stepped care system for depression with adults.

There are 9 randomised controlled studies, 1 quasi-randomised comparison study, and 3 uncontrolled prospective cohort studies. The number of patients per study range from 18-7859, (mdn = 430).

Recovery rates for depression are between 50%-60% in stepped care, and “equivalence to usual care is suggested by comparison studies.”, however the

evidence in some studies suggesting the superiority of stepped care, the authors find to be inconclusive.

Analysis

This systematic review positively met all the criteria of the PRISMA checklist apart from there was no information regarding any protocol and registration or funding.

The authors use a published quality checklist which is suitable for RCT's and non-randomised studies, they did not rate any of the studies as excellent, and found all the RCTs to be good quality, of the non-randomised studies two were found to be fair, and two poor quality. The authors discuss and critique the variety of quality, underpowered studies and missing data, lack of comparators, and variety in results reporting and acknowledge that this made "comparisons or meta-analysis more difficult." The authors also acknowledge that the inclusion of non-randomised controlled trials could be seen as inferior to a review of RCT's only, however justify the inclusion "in order to gather a wide, evidence base and to enable to realities of clinical practice to be closely reflected."

The studies included are also those that focus on comorbidity with physical conditions, e.g depression and cancer or diabetes, and also use of pharmacotherapy.

The authors differentiate service delivery models by defining stepped care models as *intervention systems*, and usual care or other care systems as *comparison systems*.

The authors discuss the demographics of the study samples, mentioning employment rates being low in those reporting, acknowledging selection biases within the studies may have influenced this, and causal factors are also considered, e.g. co-morbid physical conditions, and socioeconomic factors. There are further variations between the studies regarding ethnicity and nationality and gender, i.e range for male participants between 0-56% between 13 studies.

Clinically a variety of different diagnostic measures are used amongst the studies, but most are considered appropriate.

Of particular interest is the analyses of two studies (20,34) that evaluate IAPT sites, the authors score these both fair in quality, however neither of these studies focus

specifically on the effectiveness of the stepped care system, although both are concerning the evaluation of IAPT pilot sites, where the stepped care system is assumed.

Those studies that undertook follow-up demonstrated mixed results, with two demonstrating significance for stepped care at 6-18 months which then varies at 24 months, however the van Straten et al's study (33) did not have the power "detect the significance of trends in favour of stepped care that were observed."

Although the authors view the 11 studies that compare stepped care with other service models as "tests of the efficacy of stepped care", there are some issues with the results in some, Seekles et al (40) is underpowered, this systematic review questions the results for depression with an 86% comorbidity rate, and it also has high attrition rates.

Although the authors methodology is sound, and they acknowledge the issues regarding the variability of study methodology making a comparison and meta-analysis difficult there is no discussion regarding the variability between all studies regarding patient complexity, and in particular the potential impact of co-morbidity on outcomes.

Unsurprisingly the authors found the studies stepped care systems to be heterogeneous, with a variety in types of treatment and timing of step up, and clarification of rationale for step up. The authors also acknowledge the difficulty in understanding effect, as confounding factors of implementation could impact on effect, for instance with a mixed model delivery of collaborative and stepped care, is difficult to ascertain the effect attribution to each element.

The author acknowledges limitations of the review from both a difficulty with the literature itself, and also some methodological weakness.

4.5 Summary of literature review

The review of the literature finds stepped care to appear effective however there are methodological issues amongst the studies within the systematic reviews, and both find a wide range of heterogeneity regarding definition and implementation of stepped care, which further raises questions, and potentially weakens study results.

The Richards et al (32) study is of particular relevance to this study, firstly with the methodology of an observational design, and secondly comparing four different sites defining and delivering their interpretation of stepped care. This is an important point, as it demonstrates the wide range of interpretation of the NICE guidelines, of which there is a commonality with all the studies within this review, and particularly regarding the heterogeneity of stepped care implementation.

The Chan et al (44) study is included because of its attempt to compare two clear steps in an IAPT service, therefore comparable to this study site, and attempts to offer analysis of clinical effectiveness of a fundamental design of a stepped care system, rather than comparing to specific psychological treatments.

As discussed there appears to be service data recording issues, a small sample (n=100) is used, and there are a number of confounding variables. One fundamental issue that impacts on the findings of this study is no explanation or consideration of ratio of step up, and it is not clear whether the service system is using a pure model of stepped care, or a matched care, or if the sample selection was a mixture or distinctly separate low and high? Whether if high they had previously received low intensity as this could impact on outcomes.

The study does find proportionally in favour of low intensity treatment in terms of clinical outcomes, although there is no statistical difference. Due to the lack of clarity of model it is difficult to directly compare results to this study however some inferential observations can be made.

The two systematic reviews have six studies in common, with all but two of these the focus is co-morbid depression with a physical condition. Methodological issues are found by both reviews, demonstrating the difficulty with measuring effectiveness of stepped care provision, even with randomised controlled trials. Both demonstrated a wide range of heterogeneity, which raises two possibilities, either the research

inclusion criteria were too wide, and this in itself widens the definition of stepped care or it demonstrates that the stepped care guidance or at least the interpretation of guidance implementation in routine practice is varied.

Firth et al (38) discuss three key areas of heterogeneity, there is a large variation of sample demographics amongst the studies, “marked variation in the effectiveness of “usual” care”, and considerable variation between the stepped care systems themselves, all of which mean that comparison and demonstration of model effectiveness is difficult, which is the conclusion of both the systematic reviews, with the generalizability limited. (39)

Both systematic reviews conclude that stepped care appears to be effective; however both raise issues regarding the comparators “usual care”. Firth et al (38) discuss lack of clarity as to whether usual care interventions are the same as those delivered in the stepped care systems, stating that it is unlikely, and therefore there is a possibility of effectiveness achieved through the treatment interventions themselves rather than the delivery model. Van Straten et al (39) discuss further, stating that four of the reviewed studies appear to have care as usual “probably more closely resembled ‘no care’.”, therefore demonstrating that stepped care was more effective than no care. This also supports Firth’s (38) notion that any effectiveness could be achieved through the actual treatments delivered rather than any influence of model delivery.

At best the literature proposes that a stepped care model is effective, however it is not clear whether the model itself makes a difference or whether effectiveness is achieved through evidence based treatments alone, regardless of model delivery. Both systematic reviews recommend future research should compare stepped care to matched care, or against high intensity treatment, where the treatment is the same in different models. Furthermore the literature demonstrates variety of interpretation of what is stepped care, thus weakening any finding in support of stepped care as a delivery model, and as van Straten et al (39) recommend there is a need to further test the effectiveness of different variations of stepped care in terms of cost, acceptability and clinical outcomes.

Chapter 5: Specific study context

Drawing the main points in previous sections together, psychotherapy is effective in treating anxiety and depression disorders, CBT in particular is recommended, however it is not superior to any other intervention, including low intensity treatments. There is an absence of clear evidence regarding the specifics of what best works for whom, which would be needed for the efficacy of a matched care model. The literature review finds that generally the model of stepped care is effective however implementation in practice and indeed in some studies show variation of definition of stepped care. Similarly guidelines appear to describe a mixed model of allocated/matched care and pure/progressive stepped care. Outstanding is a need to compare more specifically matched care and pure stepped care.

This study aims to explore the question regarding effectiveness in routine practice of different models of stepped care and if a service design impacts on clinical outcomes.

5.1 The study site

The study site is a large IAPT service covering the areas of Durham and Darlington, a large geographical area combined of rural and suburban, small villages and larger towns, with both areas of affluence and deprivation. The Census 2011 population count for persons aged 16+ living in County Durham Darlington was: 510,615. It is estimated that 65,561 of this population will have a mental health problem. The National Key Performance indicators for IAPT require IAPT sites to have 15% of the population with common mental health problems to enter treatment. (46) For the study site this is 9,834 per year.

This study site was a 'third wave' site meaning it was part of year 3 of the national roll out of IAPT sites. This has meant the service has benefited from lessons learnt from previous waves, however has had shorter time to develop and embed as a new service before increases in national performance targets, such as the move in 2013 from 45% to 50% recovery rate key performance indicator. This may have impacted on progress and target achievement of the service.

5.2 Author's involvement

The author is currently the Clinical Lead for the study site, and has also previously worked as a senior therapist in another IAPT site. The role of Clinical Lead is primarily to provide clinical governance, linked with improving clinical performance which includes exploring and implementing those evidenced based components which may improve a patients experience and ultimately clinical outcome. The author is also involved in a regional forum where learning and practice is shared between the IAPT services across the North East region. From personal observations it would appear that there is some variation across the North East of England regarding definition and implementation of stepped care, and it would therefore be reasonable to assume that this variation is replicated nationally.

5.3 Staffing profile of service

The study site was organised through a combination of configuration of the previous primary care mental health teams and recruitment on a large scale for trainees. Some of the previous primary care staff was trained as Psychological Wellbeing Practitioners, (PWPs) and high intensity therapists in the year prior to local roll out, to ensure there was a sufficient number of qualified therapists. None the less during the first year of operation, approximately two thirds of the workforce were in training, which naturally impacted on throughput and recovery rates, with trainees operating with lower caseloads due to university attendance and demands, and as learning how to delivery therapy, it is natural to expect that their recovery rates of their cases would improve over time alongside their competence levels. "The trainees' knowledge of CBT, ability to conceptualize, and actual therapy skills significantly improved over time." (47) The numbers of trainees within the workforce decreased over years 2 and 3, however year 2 saw some trainees taking longer to complete their training than was predicted. Subsequently the two local University training providers have lengthened their High Intensity training course from 12 to 18 months. Given the potential propensity of any improvement in clinical outcomes being affected by skill and competence development over time, this condition will need to be considered.

With a large service it is a natural occurrence to experience a turnover of staff, which did occur over the same time period. This could impact on turnover and outcomes with loss of skilled therapists, recruitment timescales causing reduced capacity, and replacement staff taking time to acclimatise to service culture and processes may affect the build-up of productivity. Replacement staff may not be at the same skill and competence level as their predecessor and again this may have an effect either way on productivity and outcomes. During the latter half of the 4 year time period analysed, the service undertook a staffing configuration. The workforce profile at commencement was 70% high intensity therapists and 30% low intensity as specified in the original service specification set out by commissioners. However with the volume of referrals at over 10,000 per year, and building waiting lists, the introduction of a progression model of stepped care re-emphasised the focus on step 2 interventions. Shorter dosage lengths of treatment compared to High Intensity was hoped to increase the volume patients treated, would improve turnover and meant that more step 2 therapists were needed. The service therefore decided to not replace like for like, and as high intensity staff left they were replaced with low intensity workers. A further variation of this was that due to the difficulty in recruiting qualified PWPs and often vacancies being too far away from the university course commencement to recruit trainee PWPs, the service created Therapy Support Workers. (TSW) These posts are trained in house to deliver only screening and remote treatment interventions (Computerised Cognitive Behavioural Therapy CCBT, and Telephone Guided Self Help TGH). Both treatment interventions are standardised computer programmes or workbook module manual guided and the function of the worker is to provide support as the patient works through the intervention, and to risk assess each week. However the TSW posts were a new job description and competence level at recruitment was more or less similar to trainee PWPs, but without the support of University training in psychological interventions.

5.4 Performance management

Historically mental health psychological therapy services have perhaps delivered therapy according to the individual clinicians training and experience, and understanding of evidence based practice informed by NICE guidelines. Increasingly there is a move towards a culture of evidenced based service delivery and cost effectiveness, meaning increased data collection analysis and monitoring, linked with funding arrangements. The national IAPT programme is possibly the largest reform of psychological therapy service delivery, and brought with it the change in culture where services are expected to demonstrate their effectiveness and are measured across a number of areas, as described previously. The KPI achievements of each IAPT service are published via a publicly accessed health database, the Health and Social Care Information Centre (HSCIC), and therefore this in itself is a performance management strategy using competition to influence potential improvement. “There is considerable evidence that the publication of provider performance measures leads to improved performance.” (48) Certainly it could be argued that some of the KPI’s place a perverse incentive to design systems a particular way or accept unsuitable patients to increase numbers. For instance, the definition of completed treatment is two or more sessions, a service could maximise its attainment of high numbers for this target by accepting unsuitable patients and ensuring they have 2 sessions. The potential negative consequence of this is that it takes up unnecessary resources and may reduce service capability of achieving the target for proportion recovered. Perhaps one of the more fundamental changes that brought performance measurement more to the forefront for the clinician was the requirement within IAPT to collect psychological measures every single session. Previously, psychological measures were traditionally collected by the clinician at the beginning, middle and end of therapy. Unfortunately when trying to monitor effectiveness, this would have resulted in a large amount of incomplete data, as it is fairly common that patients do not necessarily receive the total dosage of treatment that is recommended by NICE, with patients not attending a last session. Administering measures every session means that this data is collected regardless of when or how the treatment ends. Anecdotally, clinicians in this service initially were concerned that administering measures every session would be detrimental for treatment, the patient would be reluctant to complete so frequently, and would perhaps refuse. In reality this has not

been the case, with the vast majority of patients happy to complete every session when it is explained that this is part of the treatment, and can help both clinician and patient to track progress. The service began to develop and implement performance management strategies. This began with explaining and showing the information about the KPI's with all staff, to encourage understanding and dialogue in team meetings about the service performance and how clinical delivery was linked with achieving targets. This strategy was informed by the principle of target achievement is everyone's responsibility. Alongside the KPI reports, using the electronic database, the service developed a number of throughput and outcome monitoring reports, in the form of scorecards for both of teams within the service and of individual clinician. These reports formed a framework of performance management, which set out the frequency of reports, how they were to be used, and also the benchmark of minimum standards regarding throughput, efficiency, engagement and outcomes that individuals and the whole service needed to achieve, in order to maximise the ability to meet the KPIs. One of the main aspects was that each clinician would receive their monthly scorecard, showing them their months performance in terms of efficiency, throughput, engagement and outcomes, and that this would be discussed with their team manager in supervision, with the aim that this would encourage personal ownership of performance, reflection and a proactive attitude to engaging in improving where needed or maintaining good performance. "Individuals will respond to performance measures in ways that maximise their own utility or benefit." (49) So the values that are triggered to improve performance may be different with each member of staff, for instance a more altruistic therapist may be motivated through the goal of seeing patients recover, whilst some may be more motivated through competition with other colleagues. Similarly to the reaction of clinicians to the introduction of the progression stepped care model, many clinicians initially found this performance management culture new and different, and for some perhaps challenging. The emphasis of performance management strategies began to occur towards the end of the first year of delivery, and steadily increased to the point the scorecards given to individual therapists by the end of year 4. Whilst this can be seen as developing at the same time as the change in model delivery, they evolved over time rather than being a whole set performance management strategies introduced at a specific time. This makes it extremely difficult to control as a variable and measure the impact. Equally there is little research that explores the

effectiveness of such within the health care field, “the kind of academic rigour that has been applied to evaluation of performance management in the training field has not yet occurred in health.” (49) Therefore it is difficult with no benchmark to even estimate the potential influence of performance management on outcomes. Given the likelihood of performance management strategies having negative or no impact for some therapist’s behaviour and outputs, it could be argued that these would balance out any potential positive effect. Therefore although possible, it is likely at best performance management strategies would have an extremely small direct causality on any improving rate of clinical recovery in this study site.

5.5 Delivery of stepped care – service variation of model

There are a number of variations of phrase that describe models of stepped care in the literature, stratified, matched/ allocated and stepped, or progression. As described previously, for the purposes of this study I will define an allocated model as that which allocates a particular treatment /step based on the assessment of the patient’s complexity and severity, including psychological measure scores. A progression model is one in which a patient will be allocated onto the lowest treatment/step and ‘progress’ upwards to receive more intensive treatments depending on need and lack of significant improvement/recovery. Locally, within this study site the ‘progression model’ of stepped care means most patients receive a step 2 intervention as their first part of treatment, regardless of severity or complexity. If at completion of this step 2 treatment the patients presentation warrants further treatment they are then stepped up to the most appropriate type of therapy the service offers, or referred on to different services.

The study site began operation in its first year with an allocated model. However began the introduction to a progression model during the latter of year 2. Initially this was to pragmatically address unacceptable waiting times, an underuse of step 2 practitioners, and to put into practice the notion that it was clinically better and safer to offer patients some initial generic treatment that may help to reduce and stabilise symptoms., rather than complex patients sit on a waiting list for some length of time with no intervention and their presentation possibly worsening. Further benefits were to identify and manage any risk at the earliest point.

Cost effectiveness is an increasingly important issue in the delivery of psychological service, and as described previously one of the compelling arguments supporting the IAPT model is the projected zero cost due to the eventual reduction of welfare provision through the improvement of mental health and the subsequent increase of people entering into paid employment. (3)

Increasing the use of step 2 interventions to treat more patients is inevitably more cost effective in comparison to the equivalent use of step 3 due to the difference in the pay scales of low and high intensity staff respectively, commencing on above £21,000 for bottom band 5, and up to £41,000 for top of band 7 for qualified staff on NHS bandings (Agenda for Change bandings 2014. (www.nhscareers.nhs.uk)). Added to the lower salary cost issue, by the nature of the difference between the dosage each intervention offers, 4-8 low intensity compared to 12-16 high intensity, resulting in higher volume and shorter throughput, means that the cost of delivery is further reduced proportionally for numbers of patients receiving step 2 interventions. However cost effectiveness should not be at the detriment of clinical outcomes, and therefore if on balance low intensity interventions can be compared and achieve similar clinical outcomes to more traditional high intensity interventions, with the same type of patients, then the cost effectiveness consideration becomes more valid.

Service model change was introduced through initial meetings with staff, to discuss the problems with underuse of step 2, and a large waiting list at step 3. It was here that the majority of staff expressed the belief that where they perceived patients to have a more complex presentation, the higher the scores on the psychological measures, those patients should be matched to a step 3 high intensity treatment. Staff understood the concern regarding leaving people on a waiting list with no treatment at all, and understood the rationale of offering people some treatment quickly may be better than nothing at all. The pure definition of stepped care was explained, and the service model introduced where patients would receive a brief step 2 low intensity intervention in the first instance, and if they had not recovered at the end then they would be stepped up to high intensity treatment. The focus of this was not about treating complexity and severity with step 2 interventions, more that they would be offered as the first part of a package of treatment, and recovery from them alone would be a positive consequence if it was achieved. The change from a

more traditional allocated stepped care model to the progression model was a culture shift for staff within the service, which appeared difficult for some. The rationale was discussed in several staff meetings, and operational policies and procedures amended to promote adherence to the changed model delivery.

The study site operates through two main hubs, where referrals, triage and telephone assessments occur. Although the service operates with standardised policies and systems across both hubs, observations and discussions with staff indicated that there appeared a difference in culture, clinical decision making, and mostly a continuation in one hub area of the allocated model, rather than progression, despite a service directive to operate a progression model. This meant proportionally more patients were being allocated straight to step 3 from assessment rather than step 2. The rationale anecdotally given by individual staff and through discussions in team meetings were that the more complex and severe the patients presentation, the more likely they would be allocated straight to step 3, despite a much longer waiting list. There was a consistent belief expressed that only a high intensity therapist with their qualification and experience would be able to achieve a positive outcome with complex and severe presentations.

However it was observed that in the hub where the progression model was operating more consistently, (hub A) waiting times were decreasing, as more patients were being seen, and monthly recovery rates were consistently improving.

Hub A also received more robust performance management of staff in comparison to the other, and therefore it could be speculated that the increase in recovery rates could be due to performance management, or the use of the progression model or a combination of both. Although there were different starting points to these service operational delivery factors, the use of the progression model took time to fully embed in hub A, and at this point both this and the performance management aspect were in operation, therefore It would be useful to know what the impact on clinical outcomes would be without one of these factors.

In the second hub (hub B) both the progression model and the performance management strategies took much longer to embed, and different starting points for both factors, however at the end of year 3 performance management strategies were also not as fully embedded as in hub A. this provides an opportunity to undertake a

comparative analysis with the data, raising a question as to whether hub A's outcomes would have been similar to hub B if the two factors were in the same stage of development, and indeed provides some useful information in relation to the research questions.

5.6 Patient presentation

IAPT was expected to improve access to a talking therapy for those patients with mild to moderate presentations, previously thought to been managed at a GP level through medication or not accessing any help. Through the ability to self-refer, it was hoped that IAPT would encourage those who didn't want to speak to their GP to access a service. Through the collection of demographic information, and self-reporting scores on psychological measures, there is now a large data set of a number of years locally and nationally which is used to report on the key performance measures, and attainment of national targets. Locally within this study site patients aged 16+ are referred or self-refer to this IAPT service who have a common mental health problem – anxiety or depression disorders, are not acutely at risk to themselves or others, and not in need of care co-ordination.

Chapter 6: Methodology

Comparison of progression and allocated service delivery models for adults with common mental health problems in a North East IAPT service.

The Project aim was to explore the impact on clinical outcomes moving from one service delivery model (allocated) to another (progression) and therefore there were two main research questions for this study.

6.1 Research questions

1. Is there a relationship between outcomes and service delivery model?

2. How does a service model impact on outcomes for severe and complex presentations?

With several sub questions that would inform the findings of the main questions:

-Is there a significant association between the outcomes on PHQ-9 and the stepped approach (allocated or progression model)?
-Is there a significant association between the outcomes on GAD 7 and the stepped approach (allocated or progression model)?
-Is there a significant association between the outcomes on PHQ-9 and any of the baseline factors such as age, gender, employment, and disability?
-Is there a significant association between the outcomes on GAD 7 and any of the baseline factors such as age, gender, employment and disability?

6.2 Ethics

A study proposal outlining the methodology, and that the retrospective data would be anonymised at source. Was submitted to Durham University ethics committee and Tees Esk and Wear Valley mental health trust Research and Development Department (R&D) for approval, discussed at the Trust's Quality Assurance group, and was subsequently agreed by all.

6.3 Background data preparation

The study site uses an N3 secure internet supported clinical database called IAPTus. It is a bespoke system that is designed specifically for IAPT services, with the background reporting set up to meet the national key performance indicators, therefore providing some level of consistency in nationally reported data. At a local level it is also bespoke with the system set up visually to replicate the care pathways and the patient journey through the organisation. It is designed to allow staff to input narrative clinical recording of each treatment session, supervision sessions, as well as psychological measures and the demographic information about patients.

At the study site, the service routinely collects and stores the psychological measures information, patient demographics and recording of clinical notes on a secure electronic data base. This is part of the IAPT national delivery and evaluation of performance against key performance indicators. Therefore 4 years of data exists already within the study site, with high levels of completion for severity ratings and outcomes which are linked to key performance indicators for the service. The psychological measures are taken at every therapy session. (22)

6.4 Data Description

The data used in this project was a sample of the routine data collected by the service as described above.

The first and last PHQ9 and GAD 7 scores of each patient whom entered and completed treatment within a 4 year timescale were extracted. Completed treatment was defined in the same way as the National IAPT KPI (patients who have received 2 treatment sessions or more) rather than the discharge reasons as defined by individual therapists, which may vary the numbers depending on definitions used. The patients and their first and last scores will be separated into bands of severity as defined by the national IAPT Data Handbook. (22)

Information can be collated and extracted from IAPTus either in raw data form to transpose onto excel or through a number of mandated and self-customised reports. As well as the data returns for the national reports on the key performance indicators, the service has a performance management framework which regularly analyses several reports to aid service and individual staff improvement. This service

information had not been analysed statistically prior, in the manner in which this study set out to do.

The service began live operation on 18/10/210. For the purposes of this study raw data over a four year period (Nov 1st 2010-31st October 2014) were extracted from IAPTus, this was done through a serious of steps given the size of the dataset and the nature of some of the extraction has to be done separately. The service received 10, 313 referrals in year 1, steadily rising to 11,573 in year 4. This resulted in 43,464 patient referred and details recorded on the clinical database over the designated time period.

Table 2: Referrals per year into service

	Year One	Year Two	Year Three	Year Four
Referrals	10313	10347	11231	11573
Dropped Out Prior to Screening	2548	2505	1987	1844
Non Engagement	2017	2099	1554	1413
Declined Treatment	84	95	148	131
Not Suitable for Service	253	203	228	263
Referred On	194	108	57	37
Dropped Out at Screening	1428	1615	1957	2002
Non Engagement	533	638	799	820
Declined Treatment	232	280	364	404
Not Suitable for Service	316	316	386	366
Referred On	347	381	408	412
Dropped Out Prior to Treatment	1189	1702	1103	1053
Dropped Out After One Session	1128	1531	1153	1327
Completed Treatment	1893	4291	5394	5145

Patients that had completed a treatment with this time period were then selected. The definition of completed treatment used for initial data extraction was the same as the national KPI's, that where a patient has received two or more treatment sessions. Whilst there may be some question regarding the validity of the treatment that is only two sessions, the actual numbers of patients this applies to is predicted to be relatively small, and those receiving a low intensity intervention which is predominately psycho-education in nature may only need this level. A further aspect

to the validity of using the KPI definition of completed treatment is that the results of this study can be compared to current and future national IAPT data, and other studies of IAPT sites.

The number of patients selected using the completed treatment KPI, and therefore the data set for this study is 16,723.

Demographic data for these patients were extracted, alongside the disorder presented at referral point, what step they entered and left treatment, the geographical locality area within the service patch the patients GP is (then grouped into the 2 hubs), number of sessions and first and last outcome measures.

Where possible and appropriate, numerical information was also grouped, for instance age was grouped into government statistics age bandings, and the psychological measures PHQ9 and GAD7 scores were grouped into the pre-defined bandings of presentation severity according to score. (22)

Data from a third measure, the work and social adjustment scale (WSAS) which reports on patient perception of how their current problems affect their functioning in several areas was also extracted to provide further outcome commentary.

The outcome information was collated in several ways; actual total score, caseness (first treatment), below caseness (last treatment), recovery, and reliable improvement, reliable deterioration and no improvement. The definition of recovery is through the use of particular self-reporting psychological scales. All IAPT sites are required to use these.

6.5 Psychological measures

The PHQ-9 is a nine question scale that measures depression symptoms frequency scoring from 0, “not at all bothered by the problem”, to 27 “bothered nearly every day”. The reliability and validity of The PHQ-9 in terms of measuring depression is good. (23) The GAD-7 is a seven question scale that measures the frequency of anxiety symptoms scoring from 0-21. The reliability and validity of the GAD-7 in terms of measuring general anxiety symptoms is good, and satisfactory with more specific disorders such as social phobia, or obsessive compulsive disorder. (22,24) The scales are used in every clinical session and the scores at the first and last sessions are used to measure recovery. IAPT data returns for the KPIs stipulate that recovery

is defined as patients scoring above clinical caseness at first session on at least one measure, and below caseness on both measures at the last session to count as recovered. (caseness = 10 on PHQ9, 8 on GAD7). Reliable improvement is defined as a reduction on PHQ-9 as equal to six points or more, and equal to 4 points or more on the GAD-7. (22,23,24) The WSAS is a five item questionnaire which measures how much the mental health problem (anxiety or depression) impacts in areas of life; work, home management, close relationships, private leisure and social leisure through a rating scale of 0-8, measuring the impairment in that area. 0 equals no impairment and 8 indicate very severe impairment. The total score measures overall functional impairment, with 0-10 showing subclinical impairment, 10-20 functional impairment; 20-30 moderate impairment and 30+ severe impairment. (22, 50) The WSAS results are only analysed in this study within the descriptive analysis, and not included within the further statistical tests as it is not used within the recovery outcome calculations within IAPT.

6.6 Cohort design

Within routine practice it is not often or practical to set up a randomised controlled study of types of service design and delivery. This study site provides a unique opportunity to observe the impact of moving from one delivery model to another. This study used the method of an observational cohort design, with retrospective data taken from routine practice in an IAPT site. Given the nature of the question related to comparing slightly different service delivery designs it would have been extremely difficult to undertake an RCT, with regards to size, commissioning arrangements, to name but a couple of aspects. As described in section 5.5, the pragmatic decision to change and improve service efficiency and organise a system so that it delivered a more pure version of stepped care provided the opportunity to explore and compare the outcomes of each method of delivery.

As described section 5.5 the service changed from a more allocated model to a progression stepped care model approximately 18 months from the commencement of the IAPT service. Although procedurally was implemented and communicated to staff at this point, it is recognised that this took time to take effect, and there was a need for several methods of communication to ensure this was fully implemented. Also, treatment lifespan can take on average 6 months at step 3 or high intensity,

therefore patients allocated straight to treatment at step 3 at the 18 month mark of service existence, may not be completed treatment until near the 2 year mark. For these reasons it is arguably acceptable to separate the cohort by the two year mark, taking years 1 and 2 as the allocated model cohort, or the baseline comparison measure, and years 3 and 4 as the progression cohort, or the experimental group.

However one of the hub areas did not fully adopt the progression model for another year, so a further method of defining the cohorts and undertaking analysis of each to compare with the first method is to separate using the North and South hubs. As described earlier, the South hub took approximately a further year to fully implement the progression model, and therefore comparing by area, would be interesting to see if there is any difference in the outcomes of each area, and also compared to the first cohort design.

One of the fundamental concerns from clinicians regarding the progression model is that patients with severe presentations would not improve without receiving a high intensity intervention from the outset. By comparing the outcomes of those scoring moderate to severe at entry point in the allocated model to the progression model will provide some observation of service design on outcome effect. Therefore to explore the above described issues, cohorts were designed as shown in the following table:

Table 3: Cohorts as defined by service delivery

Cohorts	Allocated delivery	Progression delivery
Whole service All 4 years data	Years 1 & 2	Years 3 & 4
North	Years 1 & 2	Years 3 & 4
south	Years 1, 2, 3	Year 4
Sensitivity group (from whole service)	Year 2	Year 4

Lastly a further cohort was isolated from years 2 and 4 to test the relationship between severity of baseline scores and outcomes. The participants scoring moderate to severe on both psychological measures in each year were grouped.

6.7 Statistical analysis

SPSS version 20 was used for all statistical analysis.

To answer the research questions initial descriptive analyses was undertaken, cross tabulations and chi square tests undertaken to explore the relationship between certain factors and the outcomes. Logistical regression was then undertaken to further analyse the predictive effect of certain factors on outcomes.

Firstly a descriptive analysis of the outcomes was performed by calculating percentages of patients at each level of the outcome by year and first/last outcome. Basic description of the outcomes and age were based on mean plus standard deviation and median plus interquartile. Cross tabulations and chi square were used to explore associations with baseline factors a series of tests on each cohort investigating the relationship between model and outcome, and also controlling baseline variables, gender, ethnicity, disability and employment status.

The value of *p* is set at 0.05 in terms of significance for all tests. The Hosmer-Lemeshow goodness of fit test is used to test model fit in logistic regression as SPSS does not give any other alternative. (51)

Primary analysis included a sub cohort isolating the participants who scored moderate to severe at entry and their outcomes, by service model. Secondary analyses of the same tests for were performed on further sub-cohorts as a sensitivity analysis. Testing years 2 and 4 only, was undertaken to remove any impact and variability on data output of the practical issues around first year system set up and the effect of adjusted and improved procedures regarding clinical database input. Year 3 was removed because of any potential effect of unclear model delivery across the whole data set due to the lower adoption of the progression model in the South (hub b).

Regression is a statistical technique used to predict or explain the relationship between independent variables and the outcome variable. It essentially creates an estimation of how a set of predictors (independent variables) affect an outcome (dependent variable), and through correlation can show the strength and direction of the association between one variable and another. Given the change from one delivery model to another was completed over a time period, the sensitivity cohort contains 'clean' data in terms of model delivery. Therefore logistical regression was used with this cohort to explore the relationship with model type, (allocated or progression), baseline factors, (gender, disability, employment and age), discharge reasons and first psychological scores (PHQ9, GAD7) on likelihood of outcome. (Recovery, reliable improvement, no change. reliable deterioration). The dependent variable – recovery outcome was turned into a binary variable, so the main logistical regression was recovered versus non-recovered.

Regressions were performed on cohorts of model type, those patients who had been treated in the allocated, and those treated in progression to explore the predictive effect of baseline factors, discharge reasons and first scores, and also the cohort of the sensitivity analysis group, (years 2 & 4), exploring the relationship between outcomes and model type, baseline factors, discharge reasons and first scores.

Further regressions were used on cohorts of each psychological measure's initial score severity group, to explore the relationship between outcomes and model type, baseline factors and discharge reasons, and also included the other psychological measures initial scores.

Chapter 7: Results

7.1 Descriptive analysis

The participant dataset consisted of 16,723 patients referred and completed treatment as described above. The demographic data was analysed by the respective cohorts, there was no overall discernible difference of participant characteristics by cohort therefore in general the descriptive analysis does not distinguish by cohort, but any difference is noted below. (Appendices 3&4)

Participants at entry level and steps

Where participants enter treatment demonstrably changes as the years progress, and as the progression model is implemented, there is an observable increase in numbers that enter step 2 initially, rising from 54% in year 1, to 88% by year 4.

Table 4: Steps at first session

	Year One		Year Two		Year Three		Year Four		Total	
Low Intensity	1020	53.90%	2993	69.80%	4207	78.00%	4535	88.10%	12755	76.30%
High Intensity	559	29.50%	992	23.10%	818	15.20%	392	7.60%	2761	16.50%
Not Stated	314	16.60%	306	7.10%	369	6.80%	218	4.20%	1207	7.20%
Total	1893	11.30%	4291	25.70%	5394	32.30%	5145	30.80%	16723	100.00%

Conversely there is a decrease in numbers entering straight into high intensity treatment, where although the number completing treatment is hugely different between years 1 and 4, the number entering straight into high intensity treatment in year 4 falls to below that of year 1. Proportionally to the total number of participants, each year, the difference is considerable, with 29.5% in year 1, enter step 3 initially, and by year 4 it is 7.6%.

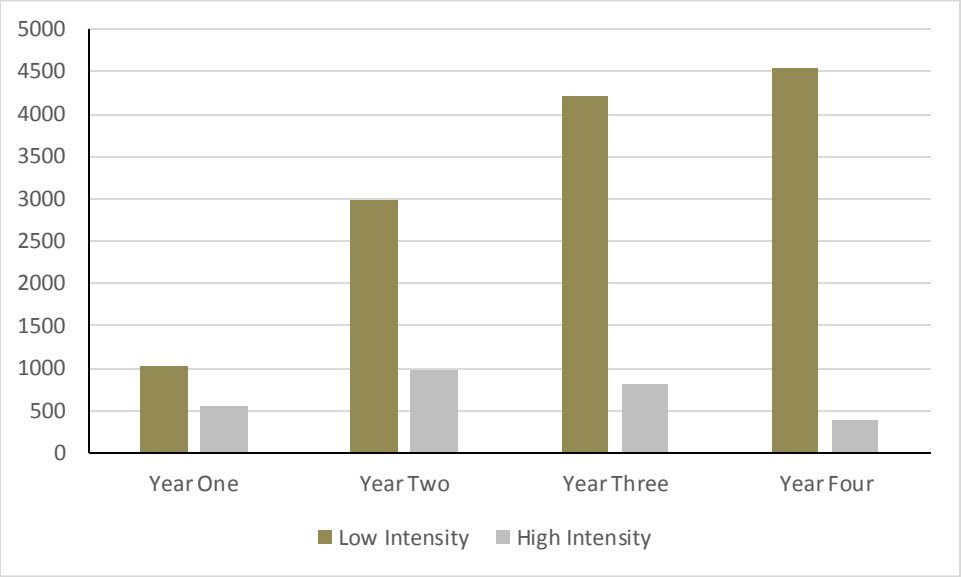


Figure 3: Step intensity at first session

Distribution of participants by entry scores

Analysis of the entry scores across the years (Appendix 5) indicates that apart from year one, severity of score distribution remains fairly static within the groups. There is no difference in the mean and standard deviation (SD) of the scores for years 2, 3, & 4.

Table 5: Entry score analysis

	Measure range	Mean	SD	Median	IQR
PHQ9	0-27	15.4	6.3	16	9
GAD7	0-21	13.9	5.1	15	8
WSAS	0-40	19.8	10.4	20	14

From year 2 onwards a mean of 59.1% of participants score moderate to severe on PHQ9, and a mean of 52% score severe on GAD7, with functioning as measured on WSAS presenting a similar picture.

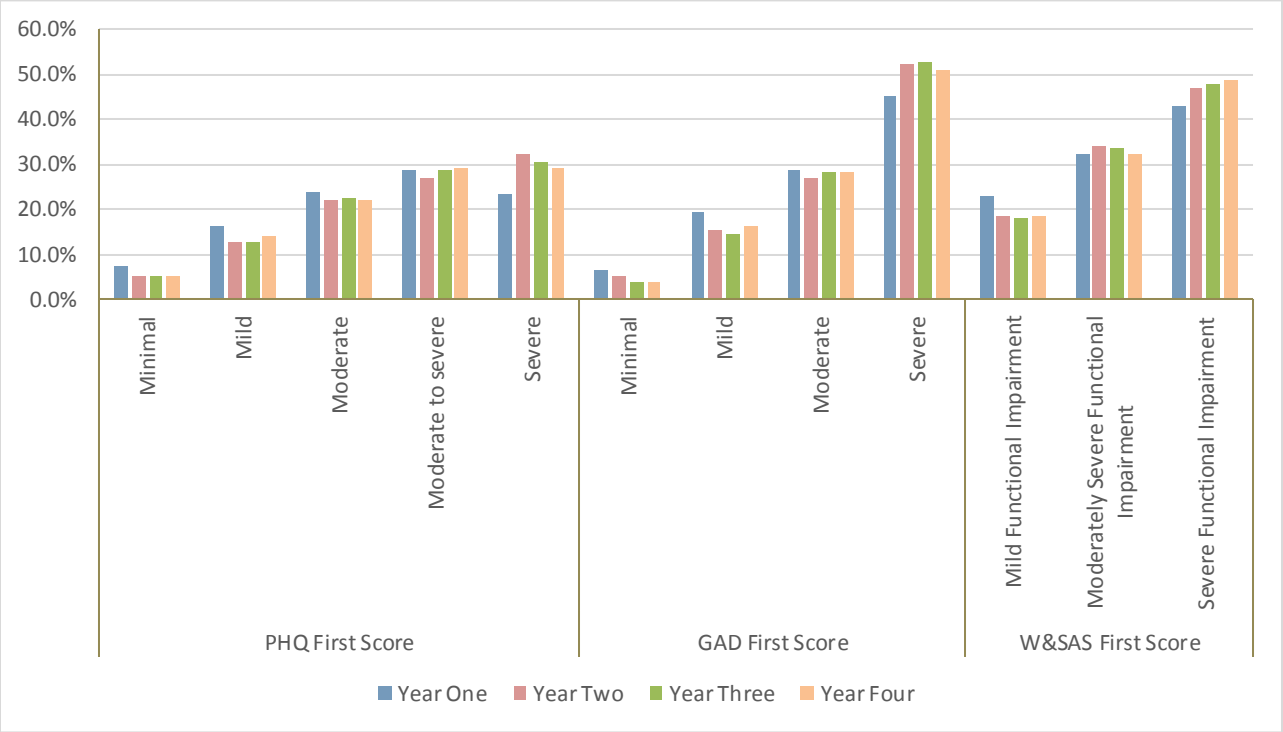


Figure 4: Scores at first session

Referred problem

A consistent pattern was observed across years 1 to 4 with the presentations of Depression, Generalised anxiety disorder (GAD), or mixed Anxiety and Depression. Depressive episode is consistent across the years with a 23% average. Other presentations occurred in much smaller numbers, and are consistent across the years with some minor fluctuations.

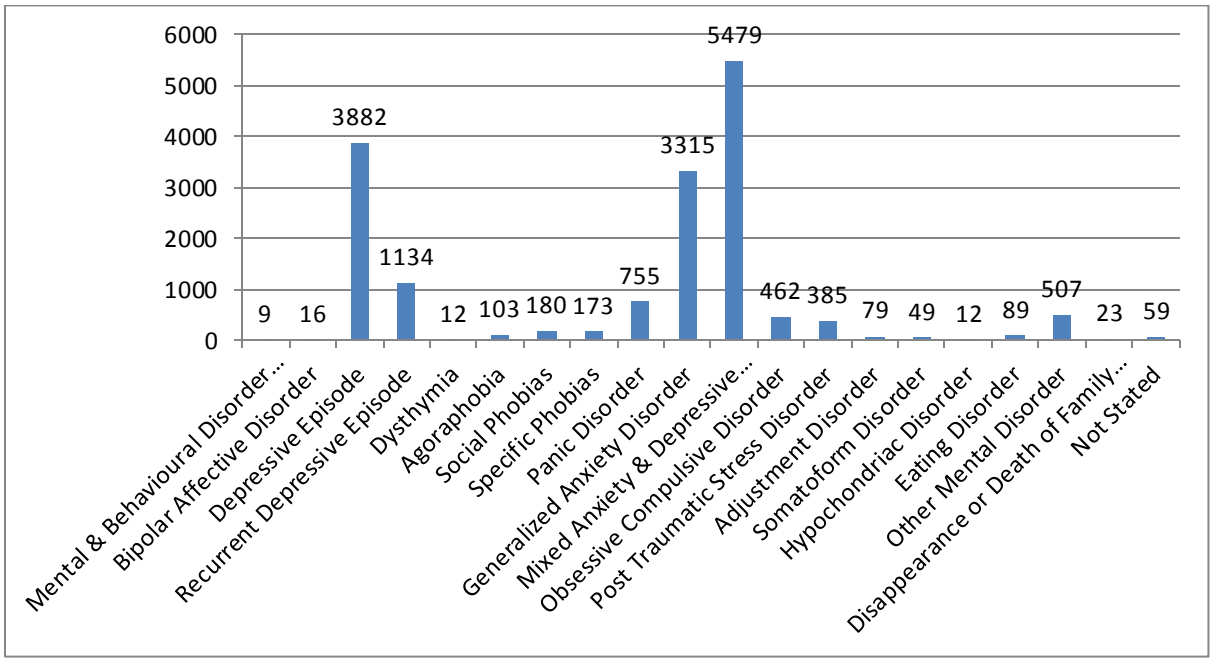


Figure 5: Referrals by disorder

With GAD there is a decrease in years 3 & 4, of 8% and 14% from years 2, conversely there is an increase for mixed anxiety and depression from years 1 – 26.8% to 39.6 % in years 4. There are some notable differences in proportion of the main 3 categories within the geographical or service model cohorts. In terms of entry scores norm and range, the only outliers are the comorbid with alcohol scoring highest on PHQ9, and specific phobias scoring lowest on GAD7 and WSAS, which is not unexpected as it is likely that the participants will be managing their anxiety through avoidance of the specific trigger of that phobia. However this data should be treated with caution given the decision regarding which label the 'problem' fits is based on a varied level of information and not always at the same stage in the pathway, and therefore no further analysis was undertaken on the referred problem due to the question regarding validity.

Ethnicity

A descriptive analysis of participants demographics (appendix 3) show that ethnicity is predominantly White British, cohort comparison does show a difference in % reported depending on model however this will be due to improved reporting in later years and therefore is not considered to necessarily show an increase in other ethnic groups given the extremely low numbers or zero in categories other than White British. (Appendix 4). There are some marginal differences in norms of entry scores

for different ethnic origin, however of particular note is how the White and Asian group score much lower than others across both PHQ9 and GAD7 but not WSAS, and the white and black African, any other mixed background and African all score much higher on PHQ9, less so on WSAS, and the only group with difference in GAD7 is the African scoring higher, and Caribbean scoring lower.

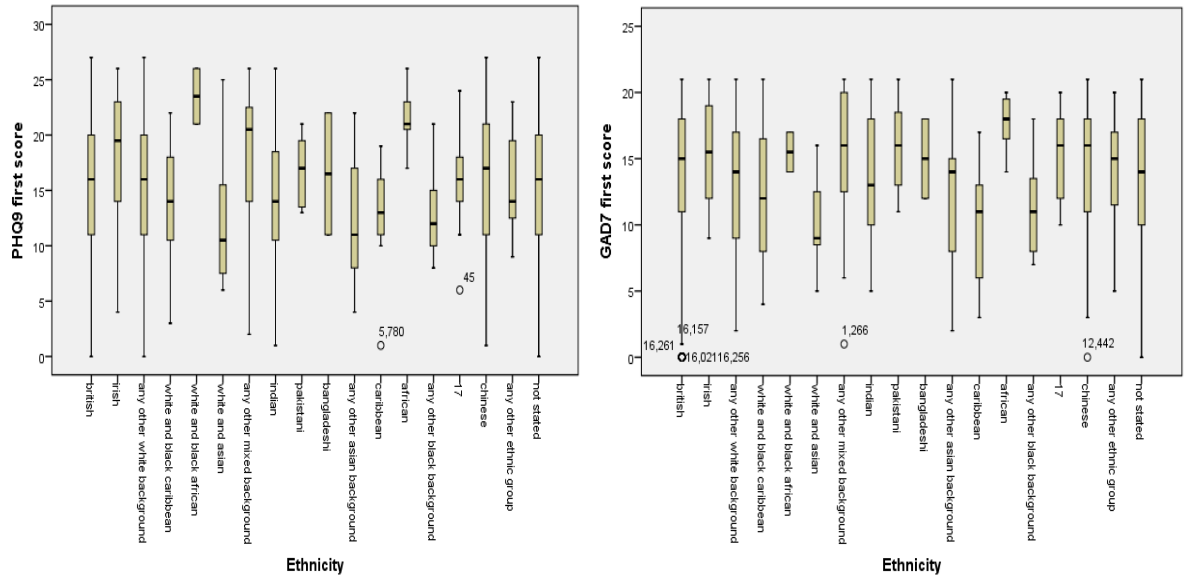


Figure 6: Ethnicity and initial scores boxplots

However numbers are very small in each of the mentioned categories and therefore cannot be generalised necessarily to that category population. No further tests were performed using the category of Ethnicity, as there would be no statistical validity.

Gender

The ratio of gender is consistent across all 4 years, even with a rise in referral and completed treatment numbers, with 36 -37% male, and 61-63% female. There is no gender difference in the norm of entry score. (Appendix 3).

Disability

There is some small variation of around 1% of those disabled in the cohort comparison; all cohorts show there are more participants who have a disability in the progression model. There is no difference with disability and GAD7 scores, but some

difference with PHQ9 and WSAS, where those registered as not disabled entry scores norm are slightly lower. (Appendices 4, 24)

Age

There is a similar consistency of age across the years, with most participants falling within one of , the 3 middle age bands – 25-34, 35-44, and 45-54, and distribution across these quite even, There are 2 notable exceptions, a decrease of 3% in year 4 in the age band 55-64, of which there is no obvious explanation, and there is a rise in the participant numbers across the age band 18-24 in years 3 and 4, by 3 % each year.

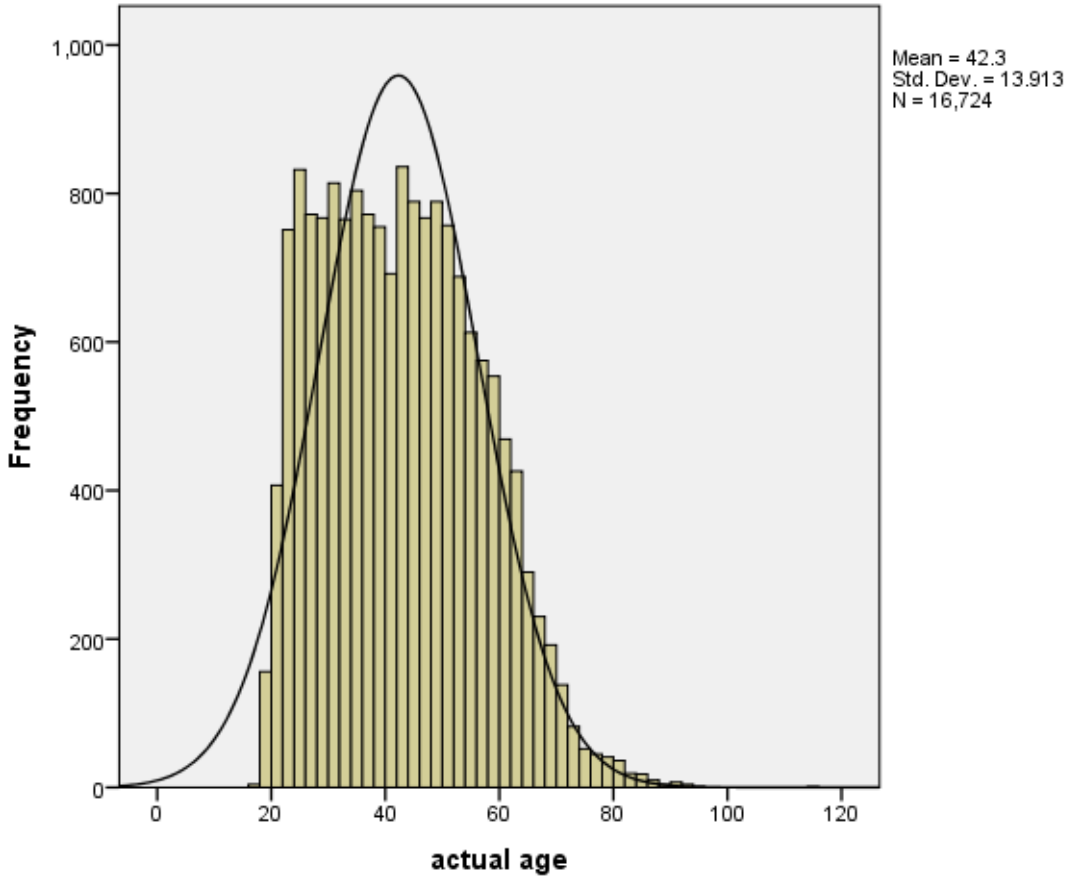


Figure 7: Referrals by age

The interquartile range is 21, with a mean age of 42.3. Standard deviation (SD) is 13.9, which further corroborates the age distribution.

Table 6: Age distribution

	Mean	SD	Median	IQR
Age	42.3	13.9	42	21

There are clear outliers with the whole group with the age groupings 16-17, and 85-94 in terms of PHQ9 scores with the range and the norm in the latter age bands showing as considerably less than middle age bands. However numbers are small, 16-17 n=3, 85-94 n=35, and therefore not necessarily representative.

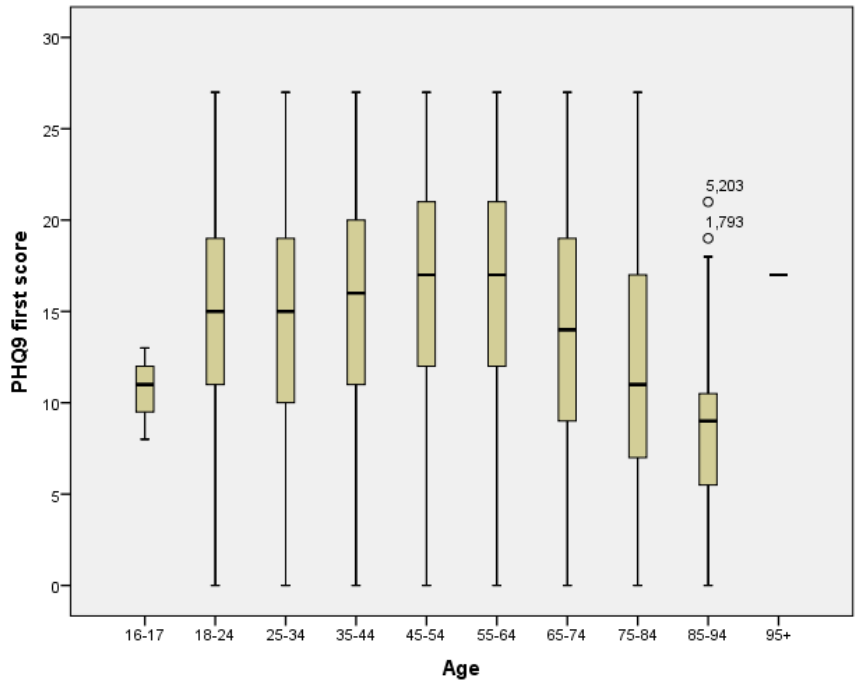


Figure 8: Age and PHQ initial scores

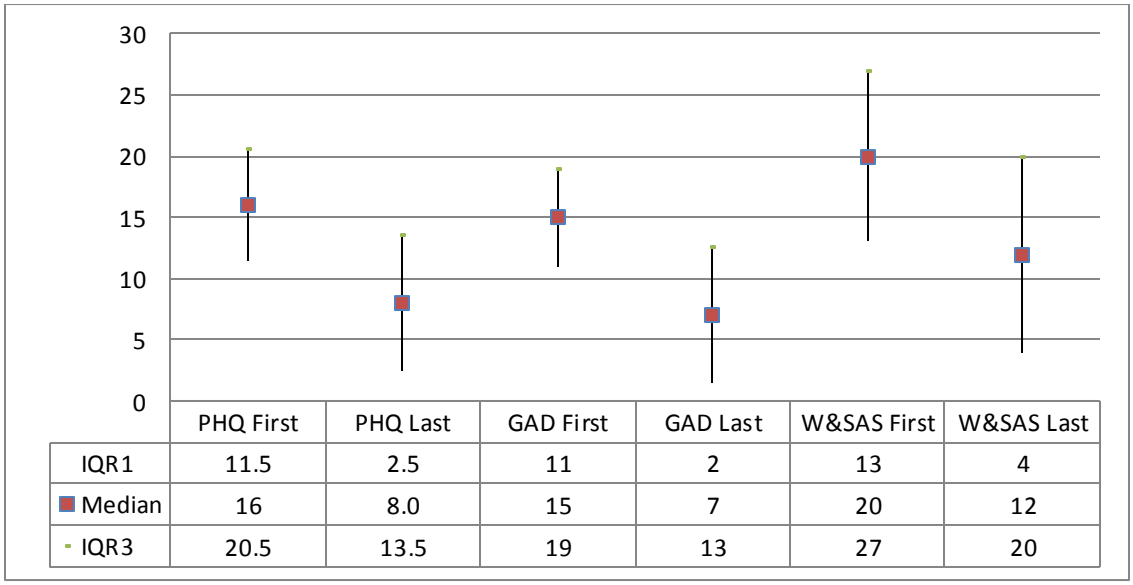
Comparing this to the other psychological measures, the score /age norm pattern on the WSAS appears similar to the PHQ9, with a drop in the norm of entry scores measuring functioning in the latter age bands; the range for the upper age bands remains similar to others. However the anxiety norm on GAD 7 remains evenly distributed, with demonstrably smaller range in scores in the age groups 16-17, and 85+, it implies a lower severity of depression and better perceived functioning in older people. However again numbers are small, and may be different if there were more participants within these age groups, and should be treated with caution. (Appendix 23).

Employment

Employment status no real difference with first score norms, apart from the retired group which scores remarkably lower, corresponding with the age distribution. There are two distinct groups where the scores in each of this group are similar. The first group contains employed, students, homemaker and retired, and the second group unemployed, sick, not claiming benefits and not working and voluntary work. The participants scores across the categories in the first group score lower than the second for PHQ9, implying that the second group experience more severe depression. (Appendix 26).

Outcomes

The spread of participants by score was calculated with the median and interquartile range for entry and outcome scores, the difference can be observed in figure 9, with a median outcome below caseness on both PHQ9 and GAD7 measures. (Appendix 6).



	PHQ First	PHQ Last	GAD First	GAD Last	W&SAS First	W&SAS Last
IQR1	11.5	2.5	11	2	13	4
Median	16	8.0	15	7	20	12
IQR3	20.5	13.5	19	13	27	20

Figure 9: Range of scores between first and last treatment session

In terms of last scores with demographic categorical factors there was no difference for norm distribution for gender, and both measures and both norms achieved below caseness. In terms of disability, non-disabled distributed with a norm below caseness for both PHQ9 and GAD7, was on the cut-off line for disabled with Gad7 and was above caseness for disabled with PHQ9, so although the norm on last

score for those registered disabled is higher than those not disabled, the norm does show a good drop in outcome scores compared to initial. For those employed, students, homemaker, no benefits or not working, or retired the norm is below caseness on PHQ9 and GAD7. For those unemployed, sick or disabled or in voluntary work, the norm for both PHQ9 and GAD7 was not below caseness in outcome scores. Ethnic groups that did not achieve norm below caseness were any other mixed, white and black Asian, African. On phq9. For gad 7 those not achieving a norm below caseness were African, any other mixed, White and black African, Caribbean, with White and Asian and Chinese on borderline. However as described previously, numbers are so low that this result should not be treated as generalizable. The distribution of last scores by age on both PHQ9 and GAD7 show that the norm falls below caseness for all age groups, with particularly lower scores in the older age groups (65+), following the pattern demonstrated with the initial scores. (Appendices 28, 29,30).

Table 7: Outcomes per year, area and model

		Year One	Year Two	Year Three	Year Four
Recovery	Whole Service	42.1%	41.2%	42.1%	50.4%
	North	44.7%	42.0%	43.0%	51.1%
	South	38.5%	39.9%	40.8%	49.7%
Reliable Improvement	Whole Service	23.0%	25.8%	25.8%	22.6%
	North	23.0%	25.2%	25.1%	21.9%
	South	23.1%	29.9%	26.7%	23.5%
No Change	Whole Service	28.3%	27.9%	27.3%	23.1%
	North	26.0%	28.1%	26.5%	22.9%
	South	31.5%	27.4%	28.3%	23.2%
Reliable Deterioration	Whole Service	6.6%	5.1%	4.8%	3.9%
	North	6.3%	4.7%	5.3%	4.1%
	South	7.0%	5.7%	4.2%	3.6%

*Values in white represent allocated model, values in grey represent the progression model.

Table 7 shows the % of participants meeting certain outcomes as defined by the national IAPT key performance indicators, by whole service and split by geographical hub, and also indicated by model type.

Recovery rates only differ slightly in the first 3 years, with a clear increase by year 4, of 8% in whole service and in both hubs. There is also a marked drop in the % of no change, and reliable deterioration, the latter particularly with the progression model.

Dosage of treatment

There appears no difference in the average dosage of treatment between cohorts, as can be observed in table 8.

Table 8: Average number of sessions per cohort

			Step 2 Only	Step 3 Only	Stepped Up	Stepped Down	Total
Whole	Allocated	Average no of Sessions	4.8	9.2	10.9	6.8	5.7
		Number of patients	58.7%	27.3%	12.3%	1.8%	100.0%
	Progressive	Average no of Sessions	4.9	10.2	12.8	6.5	7.1
		Number of patients	64.2%	10.8%	23.4%	1.6%	100.0%
North	Allocated	Average no of Sessions	4.8	9.5	10.6	6.4	6.3
		Number of patients	63.5%	23.2%	12.3%	0.9%	100.0%
	Progressive	Average no of Sessions	4.8	10	13	6	7.2
		Number of patients	64.0%	9.1%	25.5%	1.4%	100.0%
South	Allocated	Average no of Sessions	4.9	9.4	11.6	7.2	6.8
		Number of patients	56.7%	25.2%	15.4%	2.7%	100.0%
	Progressive	Average no of Sessions	4.8	10.7	12.7	6.7	7
		Number of patients	67.3%	8.8%	22.8%	1.1%	100.0%

Of particular note is the number of participants in the stepped up category, where the dosage increases by 2 sessions in the progression model compared to the allocated in the north area, and by 1 session in the south area.

Discharge reasons comparison

It can be observed that as the rate of completed treatment significantly rises, the dropout rate is also reduced, settling at consistently in years 3 and 4 at 19.5%. All cohorts show more participants completed treatment and less drop out in the progression model, although the difference in the South area is only 1%, compared to 6% in the North. (Appendix 7).

Relationship between categorical variables - Cross tabulations and chi square

Table 9 provides a summary of the chi square results of the baseline factors and outcomes, which are presented in further detail below.

Table 9: Chi square summary results of each cohort group. Baseline variables and recovery outcomes

		Whole Cohort			North Cohort			South Cohort		
		Count	Chi Square	Sig	Count	Chi Square	Sig	Count	Chi Square	Sig
Age	Recovered	6809	64.164	p<0.001	3958	41.811	p<0.001	2809	19.199	p=0.014
	Reliable Improvement	3761	37.743	p<0.001	2090	25.422	p=0.001	1646	19.117	p=0.008
	Non Recovered	4028	41.98	p<0.001	2251	22.495	p=0.004	1741	31.442	p<0.001
	Reliable Deterioration	737	15.174	p=0.019	428	6.56	p=0.255	304	3.277	p=0.773
Gender	Recovered	6808	0.893	p=0.345	3958	1.279	p=0.258	2808	4.591	p=0.032
	Reliable Improvement	3760	0.967	p=0.325	2089	0.065	p=0.799	1646	1.144	p=0.285
	Non Recovered	4026	0.015	p=0.903	2250	0.304	p=0.581	1740	0.011	p=0.917
	Reliable Deterioration	737	0.387	p=0.534	428	0.01	p=0.920	304	0.312	p=0.577
Disability	Recovered	6809	54.053	p<0.001	3958	50.334	p<0.001	2809	7.086	p=0.008
	Reliable Improvement	3761	64.726	p<0.001	2090	58.854	p<0.001	1646	1.482	p=0.224
	Non Recovered	4028	63.683	p<0.001	2251	37.839	p<0.001	1741	6.611	p=0.010
	Reliable Deterioration	737	10.686	p=0.001	428	9.089	p=0.003	304	0.872	p=0.350
First Employment Status	Recovered	6774	58.27	p<0.001	3933	51.08	p<0.001	2799	20.903	p=0.002
	Reliable Improvement	3743	28.512	p<0.001	2081	29.929	p<0.001	1637	2.092	p=0.911
	Non Recovered	4002	42.227	p<0.001	2238	39.213	p<0.001	1728	16.447	p=0.012
	Reliable Deterioration	726	118.435	p=0.428	423	3.131	p=0.680	298	1.091	p=0.955

Cross tabulations and chi square tests on age and gender showed that there is an association between age and gender; X^2 (9, n=16,718)=130, p= < 0.001, with the Chi square test meeting the assumptions of 20% or less having an expected count less than 5. The younger aged bands show significantly in favour of women, with a particular change between the proportional differences of gender in the age bands 45-54, 55-64, returning to the previous margins until the age band 85-94 where the proportion of gender is almost equal, as is demonstrated in table 10. (Appendices 11,12).

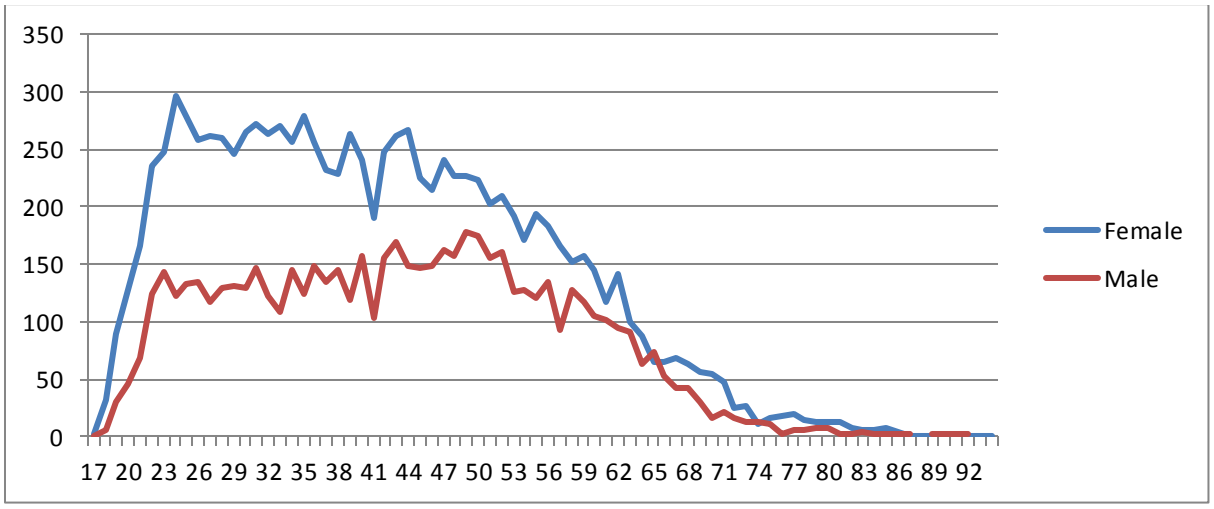


Figure 10: Age distribution by gender

Cross tabulations and chi square tests on age and disability showed that there is an association between age and disability; X^2 (9, n=16,723) = 486, *p*=<0.001, however the chi square test did not meet the assumption of less than 20% having an expected count less than 5, with an actual of 25%. However these are in the age categories 16 -17 and 95+ with the participants n=4, therefore unlikely to affect the significance. There is an observable rise in proportion of those disabled by age group which appears to correspond with the general age distribution, i.e. a peak in the age categories 45-54, and 55-64. (Appendices 15,16).

Cross tabulations and chi square tests on age and employment showed that there is an association between age and employment; X^2 (63, n=16,613) =14520, *p*=<0.001, however the chi square test did not meet the assumption of less than 20% having an expected count less than 5, with an actual of 40%. Apart from the perhaps questionable number of participants in the age groups 25-34, 35-44, that fall into the retired category, n=5, zero value in certain categories would be expected, i.e. the zero count in the range of categories other than retired for the 75+ age groups. (Appendices 13,14).

The results show that in the first two age groups the larger proportion are students, the middle groups there are more employed and the older age groups retired, all which would be expected.

Cross tabulations and chi square tests on gender and disability showed that there is an association between gender and disability; X^2 (1,n=16,718) =31.7, *p*=<0.001, with the Chi square test meeting the assumptions of 20% or less having an expected count less than 5. Proportionally to the total number of men, more males register disabled compared to the proportion of women. (Appendices 17,18).

Cross tabulations and chi square tests on gender and employment showed that there is an association between gender and employment; X^2 (7, n=16,608) = 552, *p*=<0.001, with the Chi square test meeting the assumptions of 20% or less having an expected count less than 5. More women than men are employed, and unemployed, although the proportion of men unemployed is larger than the proportion of women, a much larger number of the student category are women at

70.8% compared to men 29.2%. a larger proportion of men than women are in the category of sick or disabled, and a huge number of homemakers are women 90.8% compared to men. (Appendices 19, 20).

Cross tabulations and chi square tests on disability and employment showed that there is an association between disability and employment; X^2 (7, n=16,631) = 1081, *p*=<0.001, with the Chi square test meeting the assumptions of 20% or less having an expected count less than 5. Proportionally of those registered with a disability the larger group are unemployed (29.1%), or sick or disabled (31.6%), or retired (16.2%). (Appendices 21, 22).

7.2 Analysis of outcomes by model

Firstly the original cohorts allocated and progression defined by years, secondly the model (allocated versus progression) by area, north then south, was analysed in terms of association between model and outcomes, association between baseline factors, model and outcomes, and logistic regression to test for any significant association. Finally a further sensitivity analysis was applied.

None recorded, missing data, and non caseness at first score were removed to avoid false skew.

Table 10 is a summary of all cohorts chi square results by outcome, showing statistical significance with across all cohorts and recovery, a weaker significance for reliable improvement across all cohorts, and no significance in the south or the sensitivity analysis cohorts for reliable deterioration.

Table 10: Chi square summary results of each cohort, by outcome

Outcomes	Whole Cohort			North Cohort			South Cohort			Sensitivity Analysis		
	Count	Chi Square	Sig	Count	Chi Square	Sig	Count	Chi Square	Sig	Count	Chi Square	Sig
All Recovery Outcomes	15335	38.552	**p<0.001**	8727	14.825	**p=0.002**	6500	56.904	**p<0.001**	8578	80.021	**p<0.001**
Recovery v Non Recovery	15335	31.279	**p<0.001**	8727	13.76	**p<0.001**	6500	53.473	**p<0.001**	8578	78.608	**p<0.001**
Reliable Improvement v Non Recovery	8526	4.394	**p=0.036**	4769	1.029	p=0.310	3691	2.511	p=0.113	4733	1.354	p=0.245
Reliable Deterioration v Non Recovery	8526	4.675	**p=0.031**	4769	0.025	p=0.874	3691	2.145	p=0.143	4733	0.275	p=0.600

Allocated v progression (whole cohort)

Cross tabulations and chi square tests on outcome group by model show that there is an association between outcome and model, X^2 (3,n=15335) =38.5, p = <0.001, with more participants attaining recovery in progression model than allocated, both in number and in proportion to the total number of participants. There is no discernible difference in reliable improvement or reliable deterioration between models. Proportionally there are more participants not recovered in allocated model. (table 11, appendix 32).

Table 11: Cross tabulation of recovery and model by whole cohort

			Model		Total
			Allocated	Progressive	
Recovery	Recovered	Count	2323	4486	6809
		% within Model	41.40%	46.10%	44.40%
	Reliable Improvement	Count	1401	2360	3761
		% within Model	25.00%	24.30%	24.50%
	Non Recovered	Count	1570	2458	4028
		% within Model	28.00%	25.30%	26.30%
	Reliable Deterioration	Count	311	426	737
		% within Model	5.50%	4.40%	4.80%
Total		Count	5605	9730	15335
		% within Model	100.00%	100.00%	100.00%

Tests were then undertaken separating the dataset as recovered versus all other outcomes by model. Cross tabulations and chi square tests show an association between outcome and model, X^2 (1,n=15,335) =31.2, p=<0.001, with larger proportion of participants recovering in the progression model, and more not recovering in allocated. (Appendices 33, 34).

The outcome group reliable improvement was then isolated in a similar way (with the recovery group omitted for this calculation). Cross tabulations and Chi square tests show an association between this outcome and model, X^2 (1, n= 8526) = 4.39, p=0.036, with a 2% difference in favour of the progression model, and just under 2% proportionally are more likely to not make improvement in the allocated model. The association between reliable improvement and model is weaker than that of recovery outcome. (Appendices 35,36). The same was undertaken for reliable deterioration, cross tabulations and Chi square tests show a weaker association between this

outcome and model with X^2 (1, n=8526) =4.67 p = 0.03, only 1.4% proportionally more reliably deteriorated in the allocated model, and 1.4% more did not recover in the progression model. (Appendices 37,38).

Analysis of score severity and outcome by model

To observe the outcomes of participants in the upper end of severity scores the variables were isolated where first scores show moderate to severe or severe on PHQ9 and moderate or severe on GAD7.

Table 12 shows the chi square summary results for each cohort group, by moderate to severe scores and outcome. Each severity groups results are detailed below.

Table 12: Chi square summary results of each cohort group, isolating moderate and severe scores, by outcomes

Severity of Levels	Whole Cohort			North Cohort			South Cohort			Sensitivity Analysis		
	Count	Chi Square	Sig	Count	Chi Square	Sig	Count	Chi Square	Sig	Count	Chi Square	Sig
Severe PHQ Levels and Recovery Outcomes	4986	17.063	**p=0.001**	2785	10.045	**p=0.018**	2169	19.927	**p<0.001**	2776	27.208	**p<0.001**
Moderate to Severe PHQ Levels and Recovery Outcomes	4754	20.649	**p<0.001**	2754	9.529	**p=0.023**	1967	21.663	**p<0.001**	2664	30.142	**p<0.001**
Severe GAD Levels and Recovery Outcomes	8567	25.034	**p<0.001**	4850	13.408	**p=0.004**	3659	32.372	**p<0.001**	4759	40.753	**p<0.001**
Moderate GAD Levels and Recovery Outcomes	4694	16.351	**p=0.001**	2682	3.388	p=0.336	1977	29.624	**p<0.001**	2639	24.841	**p<0.001**

Cross tabulations and Chi square tests were undertaken on the higher PHQ9 groups and show an association between score severity, outcome and model. Severe PHQ9 showed X^2 (3, n=4986) = 17.0 p = <0.001, with participants scoring severe a larger proportion recovered in the progression model, a larger proportion did not recover in the allocated model, and there was no discernible difference with reliable improvement or reliable deterioration. Moderate/severe PHQ9 showed X^2 (3, N= 4754) =20.7, p=<0.001, with participants scoring moderate to severe a larger proportion recovered in the progression model, however a larger proportion attained reliable improvement or did not recover in the allocated model, and no real difference with reliable deterioration. (Appendices 39,40).

Cross tabulations and chi square tests were undertaken on the higher GAD7 score groups and showed an association between severity, outcome and model. The GAD7 severe group showed X^2 (3, n=8567) = 25.0, *p*=0.001, with the severe group the largest proportion to recover was in progression than allocated, and conversely a larger proportion not recovering in allocated than progression. The margins of proportional difference with reliable improvement and reliable deterioration were 1% or less. The Moderate GAD 7 group showed significance with X^2 (3,n=4694) =16.3, *p*=<0.001, with the largest proportion of participants recovered in the progression model, and less than 1% difference between models for reliable improvement or reliable deterioration. (Appendices 41,42).

Analysis of outcomes and drop outs

The data was isolated in terms of reasons for discharge from the service, as stipulated by the individual clinician to observe any difference in those 'dropping out' of treatment earlier than scheduled, by model.

There was a correlation in result with the discharge reason completed treatment and outcome, showing an association with model, with again around a 5% proportional difference in favour of the progression model for those recovered. There is a difference between the KPI defined completed treatment outcomes and therapist defined discharge reasons. (Appendix 7).

The group 'dropped out' showed some association with X^2 (3, n=3342) =12.3, *p*=0.006, with those in the allocated the largest proportion to have recovered despite dropping out, yet those in the progression model and attained reliable improvement were the largest proportion to have dropped out. There was no difference in the dropout rates by model for those not recovering or deteriorating. (Appendices 132, 133).

Analysis of baseline factors, outcomes and model for whole service data

Cross tabulations and chi square tests were undertaken on baseline factors, outcome groups and model.

Age

There is an association between outcomes, age and model, those recovered showed X^2 (9, n=6809) = 64.1, *p*=<0.001, where the age groups 18-24, and 35-44 the larger

proportion to recover were in the progression model, and 45-54 in the allocated model. Those that reliably improved showed X^2 (7, n=3761) =37.7, *p*=<0.001, where again the 18-24 group the larger proportion to improve was in the progression model, however in the 35-44, and 45-54 groups proportionally the largest to improve were in the allocated model, with there being no difference in other age groups. The association for non-recovered showed X^2 (8,n=4028)=41.9, *p*=<0.001, where in the 18-24 group the largest proportion to not recover was in progression model, however in the 25-34, and 55-64 groups the largest proportion to not recover were in the allocated model. The association is a weaker one for reliable deterioration with X^2(6,n=737)=15.1, *p*=0.019 , where reliable deterioration occur in a larger proportion in the progression model for 18-24s, but for the next 2 age groups it occurs in the allocated model, with no difference from 45+. (Appendices 43, 44).

Gender

There is no association and no difference in the cross tabulations for gender and outcomes, although cross tabulations show a reliable deterioration is proportionally larger for males in the allocated model and for females in the progression model, *p*= >0.5, therefore there is no association. (Appendices 45, 46).

Disability

There is an association between all outcomes, disability and model. For those recovered with X^2 (1,n=6809)=54.0, *p*=<0.001, reliable improvement X^2(1,n=3761)=64.7, *p*=<0.001, non-recovered X^2 (1,n=4028)-63.6 *p*=<0.001, and reliable deterioration X^2 (1,n=737)=10.6 *p*=0.001. Proportionally more registered disabled achieve all outcomes in the progression model than the allocated. (Appendices 47, 48)

Employment

There is an association between all outcomes and model except reliable deterioration. For those recovered, X^2 (7,n=6774)=58.2, *p*=<0.001, with no difference in all categories apart from a larger proportion of unemployed in the allocated model, and a larger proportion of sick or disabled in progression. For reliable improvement, X^2(7,n-3743)=28.5 *p*=<0.001, with homemaker, employed and unemployed all having larger proportion in the allocated, but again sick or disabled have a larger proportion in the progression model. For non-recovered, X^2

(5,n=726)=4.90, *p*=0.428 where no difference by model for some categories apart from a larger proportion for unemployed in the allocated model, and for students and sick or disabled a larger proportion in the progression model. (Appendices 49, 50).

North allocated v north progression

Cross tabulations and chi square tests on outcome group by model show that there is an association between outcome and model, X^2 (3,n=8727) =14.8, *p* = <0.05, with more participants attaining recovery in progression model than allocated, both in number and in proportion (4%) to the total number of participants, There is no discernible difference in reliable improvement or reliable deterioration between models. Proportionally there are more participants not recovered in allocated model. (Appendices 51, 52).

Tests were then undertaken separating the dataset as recovered versus all other outcomes by model. Cross tabulations and chi square tests show an association between outcome and model, X^2 (1,n=8727) =13.7, *p*=<0.001, with larger proportion of participants recovering in the progression model, and more not recovering in allocated. No correlation and no proportional difference between models were found for neither reliable improvement nor reliable deterioration. (Appendices 53, 54).

Analysis of score severity and outcome by model

To observe the outcomes of participants in the upper end of severity scores the variables were isolated where first scores show moderate to severe or severe on PHQ9 and moderate or severe on GAD7.

Cross tabulations and Chi square tests were undertaken on the higher PHQ9 groups and show an association between score severity, outcome and model. Severe PHQ9 showed X^2 (3, n=2785) = 10.0 *p* = <0.05, with participants scoring severe a larger proportion recovered in the progression model, (difference of 5% between models), a larger proportion did not recover in the allocated model, (again difference of around 5% between models) and there was no discernible difference with reliable improvement or reliable deterioration. Moderate/severe PHQ9 showed X^2 (3, n=2754) =9.52, *p*=<0.05, with participants scoring moderate to severe a larger proportion recovered in the progression model, (6% difference between models) however a larger proportion attained reliable improvement or did not recover in the

allocated model, and no real difference with reliable deterioration. (Appendices 59,60).

Cross tabulations and Chi square tests were undertaken on the higher GAD7 score groups and showed an association between severity, outcome and model. The GAD7 severe group showed X^2 (3, n=4850) = 13.4, *p*=0.05, with the severe group the largest proportion to recover was in progression than allocated (5% difference between models), and conversely a larger proportion either reliably improving or not recovering in allocated than progression. Reliable deterioration was less 1% proportional differences. The Moderate GAD 7 group showed no significant correlation, although there were proportional differences of 3% more recovered with the progression model and the same difference with more non-recovered with the allocated model. (Appendices 61,62).

Analysis of outcomes and drop outs

The data was isolated in terms of reasons for discharge from the service, as stipulated by the individual clinician to observe any difference in those 'dropping out' of treatment earlier than scheduled, by model.

There was a correlation in result with the discharge reason completed treatment and outcome, showing an association with model, with again around a 4% proportional difference in favour of the progression model for those recovered. There is a difference between the KPI defined completed treatment outcomes and therapist defined discharge reasons.

The group 'dropped out' showed some association with X^2 (3, n=1965) =12.4, *p*=<0.05, with those in the allocated the largest proportion (about 5% difference between models) to have recovered despite dropping out, with a larger proportion of non-recovered in the progression model. (Appendices 134, 135).

Analysis of baseline factors, outcomes and model

Cross tabulations and chi square tests were undertaken on baseline factors, outcome groups and model.

Age

There is an association between outcomes, age and model, those recovered showed X^2 (9, n=3958) = 41.8, *p*=<0.001, where the age groups 18-24, the larger proportion to recover were in the progression model, and 45-54 in the allocated model. Those that reliably improved showed X^2 (7, n=2090) =25.4, *p*=<0.001, where again the 18-24 group the larger proportion to improve was in the progression model as well as the 25 -34, however in the 35-44, and 45-54 groups proportionally the largest to improve were in the allocated model, with there being no difference in other age groups. There was a very weak association for non-recovered, and no correlation for reliable deterioration. (Appendices 63, 64).

Gender

There is no association and no difference over 2% in the cross tabulations for gender and outcomes. (Appendices 65, 66).

Disability

There is an association between all outcomes, disability and model. For those

Recovered with X^2(1,n=3958)=50.3,*p*=<0.001, reliable improvement X^2(1,n=2090)=58.8, *p*=<0.001, non-recovered X^2 (1,n=2251)37.8 *p*=<0.001, and reliable deterioration a weaker association X^2 (1,n=428)=9.08 *p*=<0.05. Proportionally more registered disabled achieve all outcomes in the progression model than the allocated. (Appendices 67, 68)

Employment

There is an association between all outcomes and model. Those recovered, X^2 (1,n=3933)=51.0, *p*=<0.001, with a larger proportion of employed and unemployed in the allocated model, and a larger proportion of sick or disabled in progression. For reliable improvement, X^2(1, n=2090)=58.8 *p*=<0.001, with unemployed having larger proportion in the allocated, however sick or disabled and students have a larger proportion in the progression model. For non-recovered, X^2 (1,n=2251)=37.8,

p=<0.001 where no difference by model for some categories apart from a larger proportion for unemployed in the allocated model, and for students and sick or disabled a larger proportion in the progression model. There is a weaker association for those reliably deteriorated, X^2 (1, n=428)=9.08, *p*=<0.05, with larger proportions of employed and unemployed in the allocated model, and larger proportions of sick or disabled in the progression model. (Appendices 69,70).

South allocated v south progression

Cross tabulations and chi square tests on outcome group by model show that there is an association between outcome and model, X^2 (3,n=6500) =56.9, *p* = <0.05, with more participants attaining recovery in progression model than allocated, with a proportional difference of 9.6% between models, proportionally more there were reliable improvement ,reliable deterioration and not recovered in allocated. (table 13, appendix 72).

Table 13: Cross tabulation of recovery and model in the south

			Model		Total
			Allocated	Progressive	
Recovery	Recovered	Count	1764	1045	2809
		% within Model	40.10%	49.70%	43.20%
	Reliable Improvement	Count	1153	493	1646
		% within Model	26.20%	23.50%	25.30%
	Non Recovered	Count	1253	488	1741
		% within Model	28.50%	23.20%	26.80%
	Reliable Deterioration	Count	228	76	304
		% within Model	5.20%	3.60%	4.70%
Total		Count	4398	2102	6500
		% within Model	100.00%	100.00%	100.00%

Tests were then undertaken separating the dataset as recovered versus all other outcomes by model. Cross tabulations and chi square tests show an association between outcome and model, X^2 (1,n=6500) =53.4, *p*=<0.001, with larger proportion of participants recovering in the progression model, with a proportional difference of over 9% and more not recovering in allocated. No correlation was found for reliable improvement nor reliable deterioration, however there was some proportional difference between models. (Appendices 73,74).

Analysis of score severity and outcome by model

To observe the outcomes of participants in the upper end of severity scores the variables were isolated where first scores show moderate to severe or severe on PHQ9 and moderate or severe on GAD7.

Cross tabulations and Chi square tests were undertaken on the higher PHQ9 groups and show an association between score severity, outcome and model. Severe PHQ9 showed X^2 (3, n=2169) = 19.9 p = <0.001, with participants scoring severe a larger proportion recovered in the progression model, (proportional difference of 6% between models), a larger proportion did not recover in the allocated model, (proportional difference of around 7% between models) and there was no discernible difference with reliable improvement or reliable deterioration. Moderate/severe PHQ9 showed X^2 (3, n=2169) =21.6, p=<0.001, with participants scoring moderate to severe a larger proportion recovered in the progression model, (10% proportional difference between models) and a larger proportion attained reliable improvement, reliable deterioration or did not recover in the allocated model.(Appendices 79,80).

Cross tabulations and Chi square tests were undertaken on the higher GAD7 score groups and showed an association between severity, outcome and model. The GAD7 severe group showed X^2 (3, n=3659) = 32.3, p=<0.001, with the severe group the largest proportion to recover was in progression than allocated (nearly 9% proportional difference between models), and conversely a larger proportion not recovering in allocated. Reliable improvement and reliable deterioration were less 1% proportional difference between models. The Moderate GAD 7 group showed X^2 (3, n=1977) =29.6, p=<0.001, with the largest proportion to recover in the progression model, with over 13% difference between models. Larger proportions were in the allocated model for reliable improvement, reliable deterioration and non-recovered. (Appendices 81,82).

Analysis of outcomes and drop outs

The data was isolated in terms of reasons for discharge from the service, as stipulated by the individual clinician to observe any difference in those 'dropping out' of treatment earlier than scheduled, by model.

There was a correlation in result with the discharge reason completed treatment and outcome, showing an association with model, with a larger proportion recovered in the progression model, and the larger proportion reliably improved and non-recovered in the allocated model. There is a difference between the KPI defined completed treatment outcomes and therapist defined discharge reasons.

There was no correlation with dropped out of treatment and outcome, however there was some proportional differences between models. (Appendices 136, 137).

Analysis of baseline factors, outcomes and model

Cross tabulations and chi square tests were undertaken on baseline factors, outcome groups and model.

Age

There is an association between outcomes, age and model, those recovered showed X^2 (8, n=2809) = 19.1, *p*=<0.05, where the age groups 18-24, 25-34 the larger proportion to recover were in the progression model, and 55-64in the allocated model. Those that reliably improved showed X^2 (7, n=1646) =19.1, *p*=<0.05, where again the 18-24 group the larger proportion to improve was in the progression model 55-64 proportionally the largest to improve were in the allocated model, with there being no difference in other age groups. Non-recovered showed X^2 (7, n=1741) =31.4, *p*=<0.001, with larger proportions in the progression model for 18-24, and 35-44, and larger proportion in the allocated model for 25-34, 45-54 and 55-64. There was no correlation for reliable deterioration. (Appendices 83, 84).

Gender

There is a weak association for recovered with X^2 (1,n=2808)=4.59,*p*=<0.05, with proportionally more females in progression and more males in allocated. There are no further associations with gender and type of outcome. (Appendices 85,86).

Disability

There is some association between some outcomes, disability and model. For those Recovered with X^2(1,n=2809)=7.08,*p*=<0.05, with proportionally more registered disabled recover in progression model. For those non recovered X^2 (1,n=1741)=6.61 *p*=<0.05, with proportionally more non recovered in the progression model. There

was no association between reliable improvement or reliable deterioration and model. (Appendices 87, 88).

Employment

There is some association between some outcomes and model. Those recovered, $X^2(6,n=2799)=20.9$, $p=<0.05$, with a larger proportion of unemployed in the allocated model, and a larger proportion of employed, homemaker and sick or disabled in progression although the proportional differences are small. For non-recovered, X^2 $(6,n=1728)=15.8$, $p=<0.05$ with larger proportions in employed, unemployed and homemaker in allocated, and a larger proportion of students and sick or disabled in the progression model. There is no association for reliable improvement, or reliable deterioration and model. (Appendices 89, 90).

Sensitivity analysis

Years 2 and 4 data was isolated to remove the possible effect on results of service development in year 1, and service model changeover in year 3, and the same tests undertaken.

Cross tabulations and chi square tests on outcome group by model show that there is an association between outcome and model, X^2 $(3,n=8578)$ $=80.0$, p = <0.001, with a larger proportion recovering in year 4, and the larger proportion for all other outcomes in year 2. (Appendices 91, 92)

Tests were then undertaken separating the dataset as recovered versus all other outcomes by model. Cross tabulations and chi square tests show an association between outcome and model, X^2 $(1,n=8578)$ $=78.6$, $p=<0.001$, with larger proportion of participants recovering in year 4, and a larger proportion not recovering in year 2. (Appendices 93,94). No correlation was found for reliable improvement or reliable deterioration. (Appendices 95, 96, 97, 98).

Analysis of score severity and outcome by model

To observe the outcomes of participants in the upper end of severity scores the variables were isolated where first scores show moderate to severe or severe on PHQ9 and moderate or severe on GAD7.

Cross tabulations and Chi square tests were undertaken on the higher PHQ9 groups and show an association between score severity, outcome and model. Severe PHQ9 showed X^2 (3, n=2776) = 27.2 p = <0.001, with participants scoring severe a larger proportion recovered in year 4, a larger proportion did not recover in year 2) and there was no discernible difference with reliable improvement or reliable deterioration. Moderate/severe PHQ9 showed X^2 (3, n=2664) =30.1, p=<0.001, with participants scoring moderate to severe a larger proportion recovered in year 4, (10% proportional difference between years) and a larger proportion attained reliable improvement, reliable deterioration or did not recover in year 2. (Appendices 99, 100).

Cross tabulations and Chi square tests were undertaken on the higher GAD7 score groups and showed an association between severity, outcome and model. The GAD7 severe group showed X^2 (3, n=4759) =40.7, p=<0.001, with the severe group the largest proportion to recover was in year 4 than allocated (over 8% proportional difference between years), and conversely a larger proportion of Reliable improvement, reliable deterioration, and not recovering in year 2. The Moderate GAD 7 group showed X^2 (3, n=2639) =24.8, p=<0.001, with the largest proportion to recover in year 4, with over 10% difference between years. Year 2 has the larger proportions for reliable improvement, reliable deterioration and non-recovered. (Appendices 101, 102).

Analysis of outcomes and drop outs

The data was isolated in terms of reasons for discharge from the service, as stipulated by the individual clinician to observe any difference in those 'dropping out' of treatment earlier than scheduled, by year.

There was a correlation in result with the discharge reason completed treatment and outcome, showing an association with model, with a larger proportion recovered in year 4, and the larger proportion in all other outcomes in year 2. There is a difference between the KPI defined completed treatment outcomes and therapist defined discharge reasons. There was no correlation with dropped out of treatment and outcome, however there was some proportional differences between models. (Appendices 138,139).

7.3 Logistic regression

A logistic regression was used with the years 2 and 4 data, (n = 8524) to firstly investigate the relationship of a number of independent variables including the type of model delivery (allocated or progression), and likelihood the outcome of recovery,

Logistic regression was also used with the years 2 and 4 data, with the cohorts grouped by model type, allocated (year 2) and progression (year 4), and the baseline factors added as independent variables to test the relationship with all outcome categories, recovery, reliable improvement, no change and reliable deterioration.

Further logistic regressions were used isolating the psychological measures score severity groups, and exploring the relationship between outcomes and the model type, and baseline factors.

Years 2 and 4 together cohort

The first logistic regression was used with years 2 and 4 data together. Table 14 shows that the data were shown to fit the model using the Hosmer and Lemeshow test ($p > .05$). The logistic regression model was statistically significant, with $X^2(25) = 3555.72$, $p < .001$. The model explained 45.6% of the variance in the recovery outcome. (Using Nagel-kerke's R^2) and correctly classified 76.5% of the cases. Sensitivity was 79.8%, specificity was 73.7%. The positive predictor value was 72.1% and the negative predictor value was 81%.

Table 14: Years 2 & 4 logistic regression table

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
	Model(1)	0.424	0.056	57.075	1	0	1.527	1.368	1.705
	Gender(1)	-0.024	0.058	0.176	1	0.675	0.976	0.872	1.093
	Disability(1)	-0.025	0.11	0.053	1	0.818	0.975	0.787	1.209
	Age			30.314	5	0			
	Age(1)	0.089	0.113	0.628	1	0.428	1.093	0.877	1.364
	Age(2)	0.227	0.117	3.772	1	0.052	1.255	0.998	1.578
	Age(3)	0.287	0.119	5.835	1	0.016	1.332	1.056	1.681
	Age(4)	0.176	0.127	1.905	1	0.167	1.192	0.929	1.53
	Age(5)	1.01	0.208	23.504	1	0	2.745	1.825	4.129
	Employment			160.801	5	0			
	Employment(1)	-0.725	0.073	99.336	1	0	0.485	0.42	0.559
	Employment(2)	-0.259	0.129	4.06	1	0.044	0.771	0.599	0.993
	Employment(3)	-0.895	0.091	96.315	1	0	0.409	0.342	0.489
	Employment(4)	-0.488	0.119	16.762	1	0	0.614	0.486	0.775
	Employment(5)	-0.73	0.181	16.297	1	0	0.482	0.338	0.687
Step 1[a]	PHQ			151.663	4	0			
	PHQ(1)	-0.159	0.224	0.502	1	0.479	0.853	0.55	1.324
	PHQ(2)	-0.661	0.211	9.815	1	0.002	0.516	0.342	0.781
	PHQ(3)	-1.034	0.21	24.188	1	0	0.356	0.235	0.537
	PHQ(4)	-1.339	0.213	39.578	1	0	0.262	0.173	0.398
	GAD			124.811	3	0			
	GAD(1)	0.057	0.244	0.055	1	0.815	1.059	0.656	1.709
	GAD(2)	-0.455	0.236	3.738	1	0.053	0.634	0.4	1.006
	GAD(3)	-0.914	0.235	15.166	1	0	0.401	0.253	0.635
	Discharge			1548.424	5	0			
	Discharge(1)	-2.774	0.085	1061.675	1	0	0.062	0.053	0.074
	Discharge(2)	-2.434	0.206	138.909	1	0	0.088	0.059	0.131
	Discharge(3)	-2.122	0.123	300.067	1	0	0.12	0.094	0.152
	Discharge(4)	-3.434	0.217	251.076	1	0	0.032	0.021	0.049
	Discharge(5)	-1.329	0.14	89.779	1	0	0.265	0.201	0.349
	Constant	2.181	0.328	44.166	1	0	8.853		

a. Variable(s) entered on step 1: Model, Gender, Disability, Age, Employment, PHQ, GAD, Discharge.

The logistic regression analysis (table 14) showed that there no statistical significance for gender or disability, however there was statistical significance for six of the predictor variables, model type, age, employment PHQ 9 initial scores, GAD 7 initial scores and discharge reason.

Patients in the progression model were more likely to recover than in the allocated model, (Wald statistic (1) = 57.075, $p < .001$, odds ratio 1.53, lower CI =1.368, upper CI =1.705).

In the age category, the age 18-24 group was the reference. Only the groups 44-54 and 65-74 showed a significant relationship with the recovery outcome, with patients aged 44-54 1.33 times more likely to recover than the group 18-24, (p=0.016). The 65-74 group were 2.75 more likely to recovery than the 18-24 group. (p=< .001).

In the employment category, employed was the reference. All groups showed significance, although the student group was an extremely weak one, (p=0.44). However the relationship for all groups was negative, therefore unemployed, students, sick and disabled, homemaker and retired were less likely to recover than employed.

With the psychological measures, the reference was the minimal group for both. For PHQ9 scores the moderate, moderate/severe and severe groups were statistically significant and all showed a negative relationship with the recovery outcome, and all were less likely than the minimal group to recover.

For GAD7 scores all groups apart from the mild group had a negative relationship however the mild group was not significant. The moderate and severe group had a negative relationship with the recovery outcome, i.e. less likely to recover than the minimal group, with statistical significance.

For both psychological measures the results show that the higher the initial score group, the less likely that patient was to recover.

For discharge reasons, completed treatment was the reference group. All discharge reasons were statistically significant, and all groups had a negative relationship with the recovery outcome, therefore only those with completed treatment as a discharge reason were likely to recover.

Allocated model

Logistic regression was used with the cohort of patients that had been treated in the allocated service model. (year 2), n= 3875. Regressions were undertaken with each outcome group as a binary, i.e. recovered versus non recovered, reliable improvement versus none, etc.

Recovered outcome (Appendix 104)

The data were shown to fit the model using the Hosmer and Lemeshow test (p > .05). The logistic regression model was statistically significant, with $X^2(24) = 146.089$, $p < .001$.The model explained 41.3% of the variance in the recovery outcome. (using Nagel-kerke's R^2) and correctly identified 75.1% of the cases outcomes. The model successfully identified 72.6% that recovered (sensitivity value), and 76.9% that did not recover specificity value). The positive predictor value was 72.76% and the negative predictor value was 76.9%.

With age the only group that showed statistical significance was the 65-74 group, with an odds ratio of 3.56 times more likely to recover than the reference group 18-24. However the confidence intervals range is rather large. (lower CI =1.979, upper CI= 6.416).

With employment, the groups unemployed, sick and disabled, homemaker and retired were statistically significant, and all less likely to recover than the reference group, employed.

The groups in the PHQ9 scores that were significant were the moderate, moderate/severe and severe, and all were less likely to recover than the reference group minimal.

The only group in the GAD7 scores that was significant was the severe, with a weak significance, p=.039, and less likely to recover than the reference group minimal.

All discharge reasons were significant, and all were less likely to recover than the reference group completed treatment.

Reliable improvement outcome (Appendix 105)

The data were shown to be a poor fit with the model using the Hosmer and Lemeshow test (p = .005). The logistic regression model was statistically significant, with $X^2(24)$ =363.745, p < .001. The model explained 13.1% of the variance in the reliable improvement outcome, (using Nagel-kerke's R^2) and correctly classified 73.8% of the cases. Sensitivity was 0.1%, specificity was 99.7%. The positive predictor value was 0.099% and the negative predictor value was 0.34%.

Only the 65-74 age group showed a weak significance (p= 0.020), and were less likely than the reference group 18-24 to attain reliable improvement.

In the PHQ9 scores category, the groups moderate/severe and severe were significant, with moderate/severe 5.09 times more likely and severe 5.42 times more likely to attain reliable improvement that the reference group minimal.

Similarly with the GAD7 category, the groups moderate and severe were significant, with moderate 6.27 times more likely and severe 12.11 times more likely to attain reliable improvement than the reference group minimal.

With the discharged reasons category only the dropped out group was significant, and 1.30 times more likely to attain reliable improvement than the reference category completed treatment.

No change outcome (Appendix 106)

The data were shown to fit the model using the Hosmer and Lemeshow test (p > .05). The logistic regression model was statistically significant, with $X^2(24)$ =661.319, p < .001. The model explained 22.6% of the variance in the no change outcome, (using Nagel-kerke's R^2) and correctly classified 75.1% of the cases. Sensitivity was 39.5%, specificity was 88.8%. The positive predictor value was 39.54% and the negative predictor value was 88.8%.

With the age category the only group to show significance was the 65-74 group, who were less likely than the reference group 18-24 to attain no change outcome.

Three groups in the unemployed category were significant, with unemployed 1.638 times more likely, sick and disabled 1.877more likely and retired 1.990 times more likely to attain a no change outcome compared to the reference group employed.

There was no significance with any of the psychological measures scores.

All discharge reasons were significant, with dropped out 5.29 time more likely, not suitable 5.06 times more likely, declined 3.95 times more likely and referred on 6.25 times more likely than the reference group completed treatment to attain a no change outcome.

Reliable deterioration outcome. (Appendix 107)

The data were shown to fit the model using the Hosmer and Lemeshow test (p > .05). The logistic regression model was statistically significant, with $X^2(24) = 273.458$, p < .001. The model explained 20.5% of the variance in the reliable deterioration outcome, (using Nagel-kerke's R^2) and correctly classified 94.9% of the cases. Sensitivity was 2%, specificity was 99.9%. The positive predictor value was 2% and the negative predictor value was 0.11%.

Two groups in the employed status category were significant, with unemployed 2.16 times more likely and sick and disabled 2.57 times more likely to attain a reliable deterioration outcome compared to those employed.

Only the moderate and severe groups in GAD7 psychological measures showed significance, with both less likely to attain reliable deterioration than the reference minimal group.

All discharge reason groups were significant, with dropped out 4.85 times more likely, not suitable 15.97 times more likely , declined treatment 5.52, 4 time more likely , and referred on 10.79 times more likely to than those completed treatment to have reliably deteriorated.

Progression model

Logistic regression was used with the cohort of patients that had been treated in the progression service model. (year 4), n=4649. Regressions were undertaken with each outcome group as a binary, i.e. recovered versus non recovered, reliable improvement versus none, etc.

Recovered outcome (Appendix 108)

The data were shown to fit the model using the Hosmer and Lemeshow test (p > .05). The logistic regression model was statistically significant, with $X^2(24)$ =2128.826, p < .001. The model explained 49% of the variance in the recovery outcome. (using Nagel-kerke's R^2) and correctly classified 78.4% of the cases. Sensitivity was 85.6%, specificity was 71.2%. The positive predictor value was 85.57% and the negative predictor value was 28.75%.

There was no significance for gender, disability or age categories.

All employment status groups were significant, with a negative relationship therefore all were less likely than the reference employed group, to attain a recovery outcome.

With the PHQ9 scores, the moderate, moderate/severe and severe group all showed significance, although the moderate group was extremely weak, p = .044. All groups had a negative relationship therefore all were less likely than the reference group minimal to attain a recovery outcome.

With the GAD7 scores the moderate and severe groups showed significance, with a negative relationship therefore both were less likely than the reference group minimal to attain the recovery outcome.

All discharge reasons showed significance, with a negative relationship, therefore all were less likely than the reference group completed treatment to attain recovery.

Reliable improvement outcome (Appendix 109)

The data were shown to fit with the model using the Hosmer and Lemeshow test (p > .05). The logistic regression model was statistically significant, with $X^2(24)$ =566.554, p < .001. The model explained 17.5% of the variance in the reliable improvement outcome,(using Nagel-kerke's R^2) and correctly classified 77.3% of the cases.

Sensitivity was 4.4%, specificity was 98.7%. The positive predictor value was 4.36% and the negative predictor value was 1.30%.

There was a weak significance (p= 0.033) for the unemployed group, with an odds ratio of 1.24 times more likely to attain reliable improvement outcome than the reference group employed.

With the PHQ9 scores, there was statistical significance with the moderate, moderate/severe and severe groups, with moderate 3.77 times more likely, moderate/severe 5.39 times more likely, and severe 7.14 times more likely to attain reliable improvement outcome compared to the reference group minimal.

Only the severe group in the GAD7 scores showed statistical significance, and was 9.03 times more likely to attain reliable improvement outcome than the minimal group.

The discharge reason category had 2 statistically significant groups, with dropped out 1.62 times more likely, and declined 1.86 times more likely to attain reliable improvement than the completed treatment group.

No change outcome (Appendix 110)

The data were shown to be a poor fit to the logistic model using the Hosmer and Lemeshow test (p = 0.041). The logistic regression model was statistically significant, with $X^2(24)$ =1051.175, p < .001. The model explained 30.6% of the variance in the no change outcome, (using Nagel-kerke's R^2) and correctly classified 78.7% of the cases. Sensitivity was 40%, specificity was 90.4%. The positive predictor value was 39.96% and the negative predictor value was 9.57%.

The age group 25 -34 showed statistical significance, with a negative relationship, where they were less likely to attain a no change outcome than the reference group 18-24.

3 employment status groups showed significance, although homemaker was weaker with p = .036. unemployed was 1.60 times more likely, sick and disabled 1.88 times more likely and homemaker 1.44 times more likely to attain a no change outcome than the employed group.

There was no significance with the PHQ9 scores, however all groups in the GAD7 category showed weak significance, with mild 3.14 times more likely, moderate 3.26 times more likely, and severe 3.16 times more likely to attain a no change outcome than the minimal group.

All discharge reasons showed statistical significance, with dropped out 9.84 times more likely, not suitable 8.95 times more likely, declined 7.61 times more likely, and referred on 12.07 times more likely, to attain a no change outcome compared to those completed treatment.

Reliable deterioration outcome. (Appendix 111)

The data were shown to fit the model using the Hosmer and Lemeshow test ($p > .05$). The logistic regression model was statistically significant, with $X^2(24) = 371.762$, $p < .001$. The model explained 27.3% of the variance in the reliable deterioration outcome, (using Nagel-kerke's R^2) and correctly classified 99.9% of the cases. Sensitivity was 3.3%, specificity was 99.9%. The positive predictor value was 4.39% and the negative predictor value was 0.13%.

Two groups within the employment status category showed statistical significance, with unemployed 2.65 times more likely, and sick and disabled 2.55 times more likely to attain a reliable deterioration outcome than those employed.

All PHQ9 groups showed significance, with a negative relationship, all groups (moderate, moderate/severe, severe) were less likely to attain a reliable deterioration outcome than those in the minimal group.

Both moderate and severe groups in the GAD7 category showed statistical significance, with a negative relationship. Both were less likely to attain a reliable deterioration outcome than those in the minimal group.

All discharge reasons showed significance, with dropped out 13.38 times more likely, not suitable 10.59 times more likely, declined 8.68 times more likely, and referred on 24.90 likely to be reliably deteriorated than those completed treatment.

Psychological measures Initial score severity - PHQ9

Logistic regressions were used with the cohort of patients that either scored moderate, moderate/severe or severe on the initial scores of PHQ9. Regressions were used for each outcome group as binary. The relationship with baseline factors, discharge reasons and GAD7 severity was also tested.

PHQ9 moderate

PHQ9 moderate – recovery outcome (Appendix 112)

With a cohort of n= 2053, the data were shown to fit the model using the Hosmer and Lemeshow test ($p > .05$). The logistic regression model was statistically significant, with $X^2(21) = 686.495$, $p < .001$. The model explained 38.3% of the variance in the recovery outcome (using Nagel-kerke's R^2), and correctly classified 76.4% of the cases. Sensitivity was 88.5%, specificity was 59.8%. The positive predictor value was 88.46% and the negative predictor value was 59.77%.

This group showed statistical significance and were 1.28 times more likely to recover in the progression model.

There was no difference with gender, although those with a disability appeared 1.19 times more likely to recover than those without, it was not statistically significant. Two age groups showed significance, with the 55-64 group 1.76 times more likely, and the 65-74 group 2.75 times more likely to recover than the reference group 18-24. Four employment status groups were significant, with unemployed, sick and disabled, homemaker and retired all less likely to recover than employed. Those with moderate or severe GAD7 scores as well as moderate PHQ9 were less likely to recover than those with additional mild GAD7 scores. All discharge reasons were less likely to recover than those completed treatment.

PHQ9 moderate – reliable improvement (Appendix 113)

The data (n= 2053) were shown to fit the model using the Hosmer and Lemeshow test ($p > .05$). The logistic regression model was statistically significant, with $X^2(21) = 256.272$, $p < .001$. The model explained 21.3% of the variance in the reliable improvement outcome (using Nagel-kerke's R^2), and correctly classified 86.53% of

the cases. Sensitivity was 1.4%, specificity was 99.8%. The positive predictor value was 1.73% and the negative predictor value was 99.77%.

There was no significance regarding model type for this outcome and appears there is no difference. There were only 3 groups in the discharge category showing significance, with dropped out 2.42 times more likely, declined 2.09 times more likely, and referred on 2.33 times more likely than the completed treatment group to attain reliable improvement.

PHQ9 moderate – no change (Appendix 114)

The data (n= 2053) were shown to fit the model using the Hosmer and Lemeshow test ($p > .05$). The logistic regression model was statistically significant, with $X^2(21) = 288.478$, $p < .001$. The model explained 20.5% of the variance in the no change outcome (using Nagel-kerke's R^2), and correctly classified 80.4% of the cases. Sensitivity was 19.1%, specificity was 96.4%. The positive predictor value was 19% and the negative predictor value was 96.4%.

This group was more likely to achieve a no change outcome in the allocated model.

The age groups 45-54 and 65-74 were less likely to have no change than the 18-24 group. The sick and disabled were 2.48 times more likely, and the retired 2.68 times more likely to not change than the employed group. The dropped out group were 6.27 times more likely, declined 4.43 times more likely and the referred on group 5.93 times more likely than the completed treatment group to not change from their initial scores.

PHQ9 moderate – reliable deterioration (Appendix 115)

The data (n= 2053) were shown to fit the model using the Hosmer and Lemeshow test ($p > .05$). The logistic regression model was statistically significant, with $X^2(21) = 210.327$, $p < .001$. The model explained 25.4% of the variance in the reliable deterioration outcome (using Nagel-kerke's R^2), and correctly classified 93.7% of the cases. Sensitivity was 6.7%, specificity was 99.7%. The positive predictor value was 6.7% and the negative predictor value was 99.73 %.

There appeared no difference in terms of model type and this outcome. The unemployed group were 2.70 times more likely, and the sick and disabled 2.81 times

more likely than the employed group to reliably deteriorate. Those with additional moderate or severe GAD7 were less likely to reliably deteriorate. The dropped out group were 9.59 times more likely, not suitable 26.96 times more likely and referred on 8.98 times more likely than the completed treatment group to reliably deteriorate.

PHQ9 moderate/severe

PHQ9 moderate/severe – recovery (Appendix 116)

The data (n= 2617) were shown to fit the model using the Hosmer and Lemeshow test ($p > .05$). The logistic regression model was statistically significant, with $X^2(21) = 911.303$, $p < .001$. The model explained 39.4% of the variance in the recovery outcome (using Nagel-kerke's R^2), and correctly classified 74.1% of the cases. Sensitivity was 83.1%, specificity was 67.1%. The positive predictor value was 83% and the negative predictor value was 67%.

The moderate/severe group were shown to have statistical significance ($p < .001$) with the model type, and 1.79 times more likely to recover in the progression model.

There was no significance for gender and disability and appeared no difference. The age group 65-74 showed significance, (p=0.014) and were 2.55 times more likely to recovery than the 18-24 group. Unemployed and sick and disabled were less likely to recover. There was no significance with additional GAD7 scores. The discharge reasons were all significant, will all less likely to recover than the completed treatment group.

PHQ9 moderate/severe – reliable improvement (Appendix 117)

The data were shown to be a poor fit with the model using the Hosmer and Lemeshow test (p =0.033). The logistic regression model was statistically significant, with $X^2(21) = 129.272$, $p < .001$. The model explained 7% of the variance in the reliable improvement outcome (using Nagel-kerke's R^2), and correctly classified 73.5% of the cases. Sensitivity was 0.1%, specificity was 100%. The positive predictor value was 0.14% and the negative predictor value was 0%.

The model type was significant, ($p < .001$), with this group less likely to reliably improve in the progression model.

With additional severe GAD7 scores there was statistical significance where they were 11.91 times more likely to reliably improve. The dropped out group were 1.62 more likely , and declined 1.49 times more likely to reliably improve than the completed treatment group, although the declined group was a weak statistical significance at p=0.44.

PHQ9 moderate/severe – no change (Appendix 118)

The data were shown to fit the model using the Hosmer and Lemeshow test (p > .05). The logistic regression model was statistically significant, with $X^2(21)$ =502095, p < .001. The model explained 26% of the variance in the no change outcome (using Nagel-kerke's R^2), and correctly classified 76.8% of the cases. Sensitivity was 35.4%, specificity was 90.2%. The positive predictor value was 35.4% and the negative predictor value was 90.2%.

There was no significance in terms of model type.

Aged group 25-34 showed a weak significance (p=0.046), and less likely to not change than the 18-24 group. Unemployed were 1.75 times more likely, sick and disabled 1.84 times more likely and home maker 1.55 times more likely (weak significance p =0.036) to not change than the employed group. All discharge reasons were significant, with all more likely to not change than the completed treatment group.

PHQ9 moderate/severe – reliable deterioration (Appendix 119)

The data were shown to be a poor fit with the model using the Hosmer and Lemeshow test (p =0.035). The logistic regression model was statistically significant, with $X^2(21)$ =185.344, p < .001. The model explained 21.5% of the variance in the reliable deterioration outcome (using Nagel-kerke's R^2), and correctly classified 95% of the cases. Sensitivity was 0.8%, specificity was 99.8%. The positive predictor value was 0.14% and the negative predictor value was 99.75%.

The model type showed a weak significance, (p=0.032) and less likely to reliably deteriorate in the progression model.

Two employment status groups showed significance, with unemployed 1.99 times more likely and sick and disabled 2.31 times more likely to reliably deteriorate than

the employed group. The additional severe GAD7 scores were significantly less likely to reliably deteriorate. The discharge reasons were all significant and all more likely to reliably deteriorate than the completed treatment group.

PHQ9 severe

Phq9 severe- recovery (Appendix 120)

The data were shown to fit the model using the Hosmer and Lemeshow test (p > .05). The logistic regression model was statistically significant, with $X^2(21)$ = 1088.246, p < .001. The model explained 44.9% of the variance in the recovery outcome (using Nagel-kerke's R^2), and correctly classified 79.4% of the cases. Sensitivity was 61.2%, specificity was 87.1%. The positive predictor value was 61.18% and the negative predictor value was 87%.

The model type was significant, (p < .001) with recovery for severe PHQ9 scores 1.606 times more likely in the progression model.

The 65-74 age group showed significance and were 3.79 times more likely to recover than the 18-24 group. Unemployed, sick and disabled, homemaker and retired were all significant and less likely to recover than the employed group. The discharge reasons were all significant and all less likely to recover than the completed treatment group.

PHQ9 severe- reliable improvement (Appendix 121)

The data were shown to fit the model using the Hosmer and Lemeshow test (p > .05). The logistic regression model was statistically significant, with $X^2(21)$ =66.702, p < .001. The model explained 3.2% of the variance in the reliable improvement outcome (using Nagel-kerke's R^2), and correctly classified 64.7% of the cases. Sensitivity was 1.2%, specificity was 99.1%. The positive predictor value was 1.19% and the negative predictor value was 99.14%.

There was no significance and no difference with model type.

The 65-74 group were less likely than 18-24 group to reliably improve, and the retired group 1.85 times more likely to reliably improve than the employed group. There was no significance with the additional GAD7 scores, and completed treated was the most likely discharge reason.

PHQ9 severe- no change (Appendix 122)

The data were shown to fit the model using the Hosmer and Lemeshow test (p > .05). The logistic regression model was statistically significant, with $X^2(21) = 737.800$, p < .001. The model explained 31.5% of the variance in the no change outcome (using Nagel-kerke's R^2), and correctly classified 73.9% of the cases. Sensitivity was 59.3%, specificity was 81.2%. The positive predictor value was 59.3% and the negative predictor value was 81.17%.

The model type was significant with no change less likely to occur with the progression model.

Those scoring severe and with a disability were 1.38 times more likely to have no change. 3 groups in employment status were significant, with unemployed 1.95 times more likely, sick and disabled 2.03 times more likely, and homemaker 1.72 times more likely than those employed to have no change. All discharge reasons were significant, with dropped out 8.04 times more likely, not suitable 8.90 times more likely, declined 6.36 times more likely and referred on 11.63 times more likely to be discharged with no change than those completed treatment.

PHQ9 severe- reliable deterioration (Appendix 123)

The data were shown to fit the model using the Hosmer and Lemeshow test (p > .05). The logistic regression model was statistically significant, with $X^2(21) = 737.800$, p < .001. The model explained 31.5% of the variance in the reliable deterioration outcome (using Nagel-kerke's R^2), and correctly classified 73.9% of the cases. Sensitivity was 1.5%, specificity was 100%. The positive predictor value was 1.53% and the negative predictor value was 100%.

The model type was not significant.

The unemployed group was 2.78 times more likely, sick and disabled 3.26 times more likely, and homemaker 3.16 times more likely to reliably deteriorate than the employed group. Only the additional GAD7 severe scores were significant with these less likely to reliably deteriorate. All discharge reasons were significant, with dropped out 9.37 times more likely, not suitable 20.68 times more likely, declined 6.04 times

more likely, and referred on 17.11 times more likely to be reliably deteriorated than the completed treatment group.

Psychological measures Initial score severity – GAD7

GAD7 moderate

GAD7 moderate – recovered (Appendix 124)

The data were shown to fit the model using the Hosmer and Lemeshow test (p > .05). The logistic regression model was statistically significant, with $X^2(22) = 92.790$, p < .001. The model explained 40.6% of the variance in the recovery outcome (using Nagel-kerke's R^2), and correctly classified 76.5% of the cases. Sensitivity was 87.1%, specificity was 63.4%. The positive predictor value was 87.1% and the negative predictor value was 63.37%.

The model type was significant (p < .001) with the GAD7 moderate group 1.49 times more likely to recover in the progression model.

The 35-44 group was 1.56 times more likely, 44-54 group 1.66 times more likely and the 65-74 group 4.03 times more likely to recover than the 18-24 group. Unemployed, sick and disabled, homemaker and retired were all significantly less likely to recover than the employed group. With additional PHQ9 scores the moderate/severe and severe group were both significantly less likely to recover. All discharge reasons were less likely to recover than the completed treatment group.

GAD7 moderate – reliable improvement (Appendix 125)

The data were shown to fit the model using the Hosmer and Lemeshow test (p > .05). The logistic regression model was statistically significant, with $X^2(22) = 158.098$, p < .001. The model explained 10.3% of the variance in the reliable improvement outcome (using Nagel-kerke's R^2), and correctly classified 84.4% of the cases. Sensitivity was 0%, specificity was 100%. The positive predictor value was 0% and the negative predictor value was 100%.

The model type was significant with those scoring moderate GAD7 less likely to reliably improve in the progression model.

With additional PHQ9 scores, the moderate group were 4.40 times, moderate/severe 9.66 times, and severe 15.11 times more likely to reliably improve. The dropped out group were 1.57 times, declined 1.77 times, and referred on 1.70 times more likely to reliably improve than the completed treatment group.

GAD7 moderate – no change (Appendix 126)

The data were shown to fit the model using the Hosmer and Lemeshow test (p > .05). The logistic regression model was statistically significant, with $X^2(22) = 397.927$, p < .001. The model explained 21.9% of the variance in the no change outcome (using Nagel-kerke's R^2), and correctly classified 78.3% of the cases. Sensitivity was 24.9%, specificity was 93.9%. The positive predictor value was 24.87% and the negative predictor value was 93.85%.

The model type was significant with a no change outcome less likely in the progression model.

All age groups were significantly less likely to have a no change outcome than the 18-24 group. Unemployed were 1.52 times, and sick and disabled 11.77 times more likely to not change compared to the employed group. Dropped out were 6.67 times, not suitable 3.48 times, declined 4.15 times and referred on 5.62 times more likely to not change than the completed treatment group.

GAD7 moderate – reliable deterioration (Appendix 127)

The data were shown to fit the model using the Hosmer and Lemeshow test (p > .05). The logistic regression model was statistically significant, with $X^2(22) = 256.609$, p < .001. The model explained 25.1% of the variance in the reliable deterioration outcome (using Nagel-kerke's R^2), and correctly classified 93.5% of the cases. Sensitivity was 1.8%, specificity was 99.8%. The positive predictor value was 1.8% and the negative predictor value was 99.83%.

The model type was not significant.

Men were less likely to reliably deteriorate than women. The Unemployed were 3.67 times, sick and disabled 4.19 times, and retired 3.90 times more likely to reliably deteriorate than the employed group. With additional PHQ9 scores, the moderate, moderate/severe and severe group were significantly less likely to reliably

deteriorate than the minimal group. The dropped out were 10.08 times, not suitable 26.48 times, declined 5.83 times and referred on 18.26 times more likely to reliably deteriorate than the completed treatment group.

GAD7 severe

GAD7 severe– recovered (Appendix 128)

The data were shown to fit the model using the Hosmer and Lemeshow test (p > .05). The logistic regression model was statistically significant, with $X^2(22)$ =1805.039, p < .001. The model explained 43% of the variance in the recovery outcome (using Nagel-kerke's R^2), and correctly classified 76.8% of the cases. Sensitivity was 73.3%, specificity was 78.7%. The positive predictor value was 73.27% and the negative predictor value was 78.74%.

The model type was significant with those in the progression model 1.55 times more likely to recover than in the allocated model.

65-74 group were 2.49 times more likely to recover than the 18-24 group. All employed groups were less likely to recover than the employed. Additional severe PHQ9 scores were less likely to recover, and all discharge reasons were less likely to recover than the completed treatment group.

GAD7 severe– reliable improvement (Appendix 129)

The data were shown to fit the model using the Hosmer and Lemeshow test (p > .05). The logistic regression model was statistically significant, with $X^2(22)$ =100.910, p < .001. The model explained 2.9% of the variance in the reliable improvement outcome (using Nagel-kerke's R^2), and correctly classified 66.5% of the cases. Sensitivity was 0.2%, specificity was 99.7%. The positive predictor value was 0.24% and the negative predictor value was 99.68%.

The model type was not significant.

65-74 group were less likely to reliably improve compared to the 18-24 group. With additional PHQ9 moderate/severe scores they were 2.71 times more likely, and the severe group 3.22 times more likely to reliably improve than the minimal PHQ9 score group. Dropped out were 1.41 times, and declined 1.32 times more likely to reliably improve than the completed treatment group.

GAD7 severe– no change (Appendix 130)

The data were shown to fit the model using the Hosmer and Lemeshow test (p > .05). The logistic regression model was statistically significant, with $X^2(22)$ =1175.417, p < .001. The model explained 31% of the variance in the no change outcome (using Nagel-kerke's R^2), and correctly classified 75.8% of the cases. Sensitivity was 51.3%, specificity was 85.6%. The positive predictor value was 51.27% and the negative predictor value was 85.58%.

The model type was significant with the progression model less likely to have a no change outcome with this group.

Unemployed were 1.78 times, sick and disabled 2.04 times, homemaker 1.81 times and retired 1.76 times more likely to not change compared to the employed group. There was no significance with PHQ9 scores. The dropped out group were 8.20 times, not suitable 8.42 times, declined 6.93 times and referred on 10.93 times more likely to not change than the completed treatment group.

GAD7 severe– reliable deterioration (Appendix 131)

The data were shown to fit the model using the Hosmer and Lemeshow test (p > .05). The logistic regression model was statistically significant, with $X^2(22)$ =197.330, p < .001. The model explained 21.5% of the variance in the reliable deterioration outcome (using Nagel-kerke's R^2), and correctly classified 97.8% of the cases. Sensitivity was 0%, specificity was 100%. The positive predictor value was 0% and the negative predictor value was 99.97%.

The model type was significant, with this group less likely to reliably deteriorate in the progression model.

The unemployed were 2.26 times more likely to reliably deteriorate than the employed. The additional severe PHQ9 group were significantly less likely to reliably deteriorate. The dropped out group were 6.48 times, not suitable 9.06 times, declined 9.62 times and referred on 12.93 times more likely to be reliably deteriorated.

7.4 Summary of results

The descriptive analysis shows a normal distribution of the clinical population of this study site, regarding demographics and psychological measures. Chi square tests show a significant association between recovery and all cohort groups, including the sensitivity analysis of isolating years 2 and 4.

Score severity impact

The association between the two psychological measures and the outcomes in the logistic regression results show that compared to moderate to severe in one and minimal in another, if there are moderate to severe scores in one and the other, the recovery outcome is less likely, interestingly the reliable deterioration outcome is also less likely but the reliable improvement outcome is more likely. Therefore confirming a clinical picture that a combined higher initial score on both measures is less likely to achieve recovery, but may reliably improve.

Allocated versus progression model

Moving to a progression model increases the dosage of treatment sessions for those stepped up, compared to an allocated model by 1-2 sessions. As the rate of completed treatment rises, the dropout rate is reduced, with a larger difference in the North between models. Basic calculations demonstrate that the recovery outcome demonstrably increases in year 4 compared to others, with a comparable decrease in reliable improvement, no change, and reliable deterioration. Of particular note, is the difference of recovery percentage in the South, between years 3 and 4, when in year 4 the progression model is fully embedded.

Table 15: Mean recovery outcome, by model and by cohort

Cohort	Recovery rate (mean %)	
	Allocated	Progression
Whole service	41.4	46.1
North	42.8	46.9
South	40.1	49.7
Sensitivity analysis (Yrs 2&4)	39.6	49.2

The above table 15 demonstrates a 4-10% mean difference in reliable recovery outcome attainment in favour of the progression model, with chi square tests demonstrating statistical significance. (p=<0.001).

The results demonstrate the progression model achieves more recovery, between 4-10%, and is statistically significant. Further analysis with logistic regression of years 2 and 4 showed that patients were 1.527 times more likely to recover in the progression model than the allocated model.

Table 16: Logistic regression of years 2 & 4 cohort recovered outcome

	B	SE	Wald	Sig.	Exp(B)	95% C.I. for EXP(B)	
						Upper	Lower
Recovered	0.424	0.056	57.075	0	1.527	1.368	1.705

Note: Allocated model is the reference, therefore figures shown are the values for the progression model compared to the allocated model

Score severity and model type

For both psychological measures the logistical regression shows that the higher the initial score group, the less likely that patient was to recover, and this applies in both the allocated and the progression model. The analysis of the end scores for moderate to severe PHQ9 and GAD7 showed that there is a 5% mean difference in favour of the progression model, with chi square tests demonstrating statistical significance. . (p=<0.001).

Table 17: Mean belowcaseness end score attainment of participants with moderate / severe entry scores and by model

Measure	Allocated	Progression
PHQ9	24.90%	29.30%
GAD7	31.20%	36.00%

According to the cross tabs and chi square tests the progression model does not adversely affect those who score moderate to severe at entry to treatment, conversely more (5%) achieve below caseness on each measure in a progression model. Further with the years 2 and 4 cohorts, the progression model achieves a mean recovery that is 10% more than allocated. Comparing the model type logistic regression results, although both models allocated or progression show that the

PHQ9 moderate to severe groups are less likely to recover than the minimal group, it likely to recover than the progression.

Table 18: Summary of logistic regression for PHQ9 severity groups and outcome

		B	SE	Wald	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Upper	Lower
Moderate	Recovered	0.244	0.111	4.850	0.280	1.277	1.027	1.587
	Reliable Improvement	-0.016	0.141	0.012	0.912	0.985	0.747	1.298
	No Change	-0.249	0.121	4.218	0.040	0.780	0.615	0.989
	Reliable Deterioration	-0.069	0.200	0.119	0.731	0.933	0.630	1.382
Moderate / Severe	Recovered	0.584	0.098	35.493	0.000	1.794	1.480	2.174
	Reliable Improvement	-0.346	0.093	13.743	0.000	0.708	0.589	0.850
	No Change	-0.128	0.105	1.486	0.223	0.880	0.717	1.081
	Reliable Deterioration	-0.440	0.205	4.619	0.032	0.644	0.431	0.962
Severe	Recovered	0.474	0.105	20.472	0.000	1.606	1.308	1.972
	Reliable Improvement	-0.054	0.082	0.431	0.512	0.948	0.808	1.112
	No Change	-0.311	0.094	10.886	0.001	0.733	0.609	0.882
	Reliable Deterioration	-0.196	0.282	0.485	0.486	0.822	0.473	1.427

Note: the reference group is the minimal score group, therefore the figures shown are the values of each severity group compared to the minimal group.

This is a similar result with the reliable improvement outcome where it appears that participants in either model are more likely to reliably improve than the minimal groups however the odds ratio is better in the progression model. A no change outcome appears more likely in the progression model, and a reliable deterioration outcome more likely in the allocated model.

The regressions using the score severity cohort groups show a clear result in terms of recovery, where moderate, moderate/severe and severe all showed significantly more likely to recover in the progression model.

The results for the GAD7 moderate to severe groups are less clear cut, they appear least likely to recover or reliably improve in the progression model when compared to the GAD7 minimal group, however the isolated score groups regressions showed that both moderate and severe groups were more likely to recover in the progression model.

Table 19: Summary of logistic regression for GAD7 severity groups and outcome

		B	SE	Wald	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Upper	Lower
Moderate	Recovered	0.400	0.099	16.165	0.000	1.491	1.227	1.812
	Reliable Improvement	-0.250	0.115	4.739	0.029	0.778	0.621	0.975
	No Change	-0.258	0.106	5.887	0.015	0.773	0.628	0.952
	Reliable Deterioration	0.049	0.180	0.075	0.784	1.051	0.738	1.496
Severe	Recovered	0.436	0.076	33.002	0.000	1.546	1.333	1.794
	Reliable Improvement	-0.101	0.064	2.504	0.114	0.904	0.798	1.024
	No Change	-0.234	0.075	9.672	0.002	0.791	0.683	0.917
	Reliable Deterioration	-0.532	0.218	5.944	0.015	0.587	0.383	0.901

Note: the reference group is the minimal score group, therefore the figures shown are the values of each severity group compared to the minimal group.

Compared to the minimal group, the moderate recovers better in the allocated model, for severe both recover however the odds ratios are higher in the allocated model. Regressions using score severity cohorts showed the moderate group less likely to reliably improve in the progression model. Also the progression model appears to be more likely to have no change, however the results are either less distinct or not significant for reliable deterioration and moderate or severe GAD7 scores.

Therefore compared to their minimal groups both PHQ9 and GAD7 moderate to severe groups are more likely to recover in an allocated model. However the regressions of the moderate to severe scores cohorts showed that PHQ9 scores and GAD7 moderate to severe groups are all more likely to recover in the progression model than the allocated model. The progression model is also more likely to achieve more reliable improvement for PHQ9 moderate to severe; however for GAD7 the likelihood for reliable improvement is better in the allocated model.

Discharge reasons

The discharge reasons analyses shows statistical significance that completing treatment is more likely to have a recovered outcome. Proportionally there is no difference between the model type and dropout rate, Although in the separate model type regressions, the odds ratios are more in the progression model for all the discharge reasons, this includes the completed treatment reason also. As there has not been a specific regression grouping discharge reasons to test the predictability of model type, it can only be said that the odds ratios are higher in the progression model for all discharge reasons. In terms of the score severity groups, the more

severe the initial scores the more likely they would be discharged as dropped out, not suitable or referred on.

Therefore a conclusion can be made that overall proportionally there is no difference in dropout rate with model type. Compared to the allocated model, the progression model achieves 10% more, and there is a statistically significant likelyhood of recovery for both measures and more reliable improvement for PHQ9 moderate to severe groups, however it appears to need to offer an average 1-2 more sessions overall to patients that are stepped up.

Baseline factors impact

Descriptive analysis shows that proportionally there are more younger women than men, with the difference in gender decreasing as the age bands rise. There is a rise in proportion of registered disabled within the 45-54, and 55-64 age bands, with a larger proportion being male.

Chi square tests show there is no significant association between cohorts, outcomes and gender, which is also confirmed by the logistic regression where there was no significance and odds ratios similar across both models. The only anomaly within this was that the GAD7 moderate group showed that men were less likely to reliably deteriorate than women. Generally, overall however, gender does not have an influence on any outcome.

There is a significant association between disability and all cohorts and outcomes, apart from the South, where there is a weaker association for recovery, and non-recovered, and no association for reliable improvement and reliable deterioration. However logistic regression showed no significance and odds ratios were similar across both models. The only anomaly was that the PHQ9 severe group with a registered disability were more likely to have a no change outcome. However in general a registered disability does not have any influence on outcomes.

Chi square tests show a significant association between all cohorts, outcomes and age, although the South has a weaker association with all outcome areas and is not significant for reliable deterioration. Regression showed that the 18-24 group does not do as well as the others, with a distinct poorer recovery than other age groups in the GAD7 moderate group. However the age group that showed significance across

all areas in logistic regressions was the 65-74 group. This group, and across all score severity groups appeared more likely to recover in the allocated model, however they are also more likely to reliably deteriorate too, and less likely to attain no change or reliable improvement in the allocated model.

Chi square tests showed there is a significant association for employment status, outcomes and all cohorts. Logistic regression showed a clear result that the employed group are more likely to recover, and appeared more so in the progression model. The logistic regression tests showed that in terms of model type the unemployed, sick or disabled, homemaker and retired appeared to do better in a progression model, but are less likely to recover. In particular the unemployed or sick or disabled, are more likely to not change or reliably deteriorate (there was only a small difference between models), and the more severe the PHQ9 and GAD7 scores were the more likely these outcomes would occur. Given the result that a registered disability does not appear to influence outcome, an assumption could be made that it is the more general self-reporting sick or perceived disability that impacts on outcomes. The retired group appeared more likely to recover, reliably improve or not change in the progression model, This is an interesting result considering the age group 65-74 are more likely to recover in the allocated model, it may be that there are more patients from the younger age groups that fall into the retired group

Chapter 8: Discussion

The two key research questions were regarding the impact of different stepped care models on outcomes, and also how moderate to severe presentations may impact on outcomes, and whether there was a difference for these in terms of model design and outcome. The results show that the progression model achieves better outcomes, consistently across the different cohorts including the sensitivity analysis, with a 10% proportional difference, and statistical significance of 1.53 times more likely for recovery than the allocated model. Furthermore, the patients with initial scores in the moderate to severe on both psychological measures are all shown with statistical significance to recover, and the moderate to severe PHQ9 groups are shown to be more likely to reliably improve, in the progression model, compared to the allocated model. One of the 'costs' of attaining better outcomes does appear to be an increase of 1-2 treatment sessions for patients stepped up in the progression model,

Subsidiary research questions were whether any baseline factors such as patient characteristics had any relationship with the outcomes. The results show that gender and registered disability do not make a difference, but employment status does have a relationship, with being employed more likely to recover, and being unemployed, sick or disabled, a homemaker or retired was associated with a less likely to recovery outcome. The only age group to show a significant relationship with recovery was the 65-74 group.

There are, as with any study, limitations with the methodology, and given the observational nature of the study, the associations demonstrated within the results do not prove causality, there are potentially confounding factors that are uncontrolled which may also have a relationship with the outcomes. A more detailed discussion of the meaning of the results is presented, with limitations of the methodology and the study, and future recommendations.

8.1 Descriptive analysis

No further analysis on the variable categories Ethnicity or referred problem was undertaken due to low numbers in categories other than White British, and a lack of confidence in the data regarding referred problem, as there was a concern this may give false meaningful results.

In terms of Ethnicity, the data is not as robust. The service was quite poor in the first few years at collating and recording this particular information as can be seen by the amount of 'not stated' in the first 2 years, subsequently decreasing considerably in the next 2 years as systems were improved to ensure this data was collected. Year 4 data regarding ethnicity gives a more accurate picture of the ethnic representation of the participants, over 93% recorded as White British, and less than 1% in each of any of the other categories, and 4% not stated. Compared to the national picture this study appears to have a much lower representation of ethnic groups other than White British, with 80.5% of the general population recorded as White British. (52) It is possible that the clinical population is more skewed where ethnic minority groups may be less likely to access services, there was an under –representation of other ethnic groups reported in the year one evaluation of IAPT (11), and certainly an analysis of IAPT data nationally indicates this may be the case with 89% of those accessing treatment recorded as White British. (53) Although the results show some interesting distribution in the entry scores, with African and white and black African scoring significantly higher on PHQ9 than others, the sample size is too small (total n=9) to be generalizable or valid for further analysis.

The 'coding of referred problem' occurs at triage stage, which is based on varied levels of information depending on whether it is a brief letter from the GP, or an initial discussion between the patient and a clinician, where it is a self-referral, of which the service has seen a steady increase. IAPT services generally collect this information, and it is collated nationally as referrals received by provisional diagnosis. The picture nationally looks very similar to this study site, with mixed anxiety and depression, depressive episode and generalised anxiety disorder being the top three codes, by significant margins. (53) The varied level of information does not give a confidence in the decision of category of referred problem and is something that would be useful for the service to also undertake after assessment, where the standardisation of the

assessment would at least provide the assurance of the same framework for analysis of information, reducing the effect of quality of information. The recent systemic review Firth et al (38) showed that there were issues in a number of studies of clarity on diagnostic criteria and relating clear outcomes to specific disorders. Clearly there is a need for both practice and research to improve identification and categorisation of disorders to be able to effectively relate this to outcome evaluation.

Gender distribution within the general population nationally and locally is similar, 49% male and 50% female. For those accessing treatment in this study, proportionally there is a consistent average of 37% male, and 63% female, the gender distribution is consistent across years, and no significant difference in cohorts. Across IAPT services the distribution is very similar with 36% male, 63% female and 1% not recorded, (53) and HSCIC state that this is much "more pronounced gender difference than has been seen in secondary mental health care"- quoting the 2012/13 Mental Health Bulletin 44% male and 56% female as an example. This seems to indicate that at a primary care level women are more likely to engage in a talking therapy than men.

The increase incrementally over the years in the numbers of the younger age groups can be explained by a targeted referral recruitment, pathway and treatment by the service of "academic wellbeing" with students at the local university. The distribution of age and gender (appendices 8,9, 11) shows an interesting pattern where of the whole sample, the greatest proportion of women completing treatment is in the younger to mid age groups (24-34 – 25.1%, 35-44 – 23.4%) is greater than men.

However this shifts in the middle age groups, with the largest proportion of men completing treatment, 24.9%, occurring in the 45-54 age group. Both the age distribution and gender /age proportion corresponds with the national IAPT picture. (53) Men are more at risk of depression and suicide, three times more than females, with the highest rate of suicide for men in the 40-44 age bracket. (54) What is not clear is whether younger men are poorer at engaging in treatment, and older men more likely to engage, or are younger men less likely to suffer anxiety or depression, or less likely to recognise the problems and seek help? There is a link with other factors and older men which increase the risk with significant life events such as divorce/separation, other loss, redundancy, unemployment. (55) Therefore it would

appear that the accumulativeness of significant life events may make older men more likely to exhibit mental health problems that are perhaps more severe that they then may access treatment.

The amount of participants classified as disabled significantly rose from 3-4% in both years one and two to just less than 10% in both years 3 and 4. This could be attributed to at the same time the service was developing and delivering specific treatment to those with long term conditions, of which some may be classed as a disability, or may be correlated with other disabling conditions. It would appear that there is an interesting proportional difference with more men likely to be registered disabled at 9.1%, compared to females at 6.7%. According to the analysis in this study there appears to be a rise in middle-aged men and registered disabled accessing treatment, therefore developing a long term condition or becoming disabled may be one of the significant life events that links with mental ill health particularly for men, or it may be they are more willing to access treatment for mental health having had to access treatment for physical health problems too. There is something of interest here to explore further in terms of outcomes. Although the results appear to indicate that gender does not impact on outcome, it may be worth exploring the relationship with multiple factors, such as gender, disability, employment impact on outcome and whether combined impact on likelihood of recovery?

Another interesting observation is the decrease in entry scores for the older age groups, on face value this could imply that this group are not as depressed or anxious as younger age groups, or It is possible that the older population are less likely to access or be referred to mental health services as they may not be as aware of the symptoms of mental health problems, or it is possible that mental ill health symptoms are more easily misinterpreted as physical symptoms, for instance memory problems through depression compared to natural memory decline with age, or shallow breathing common with anxiety, difficult to distinguish from asthma or chronic obstructive pulmonary disease (COPD). Alongside the likelihood of a physical long term condition increases with age, and the figures here demonstrate that this is the case, with the proportion of those registered disabled by age category rising in the older age groups to 19.3% of those aged 75 -84, and 40% of those aged 85-94 (Appendix 15), all may be contributing to an under representation of the older

population accessing psychological therapy. Lastly there is a caution about what the data may imply for this age group given the sample size of n=181 in 75-84, and n=35 in 85-94 across all 4 years and therefore is it generalizable to the whole population in this age group, or in this age group with mental health problems, as it is likely that there are significantly more older people suffering from anxiety or depression than are presenting to this service. Certainly the findings are consistent with the national statistics of referrals of 65+ to IAPT services being lower than expected. (56) There is a question here regarding the potential reasons for low engagement, and implications for service improvement. Is it a cultural generational perception and /or misinterpretation of symptoms by patients and professionals that prevents older people from accessing psychological treatment, and how can services reach out and change those perceptions and misinterpretations, thus removing the barrier to engagement? Certainly a recent health survey (57) regarding attitudes to mental illness found that attitudes and tolerance varies with age, where "the most prejudiced attitudes were held by participants aged 65 and over." If there is a perception that those with a mental illness are 'weak' then it is likely that this will also apply to self, and that seeking help risks the same judgment from others, then this is likely to be a barrier to engaging in seeking and receiving treatment. It is interesting that the 65+ groups have good recovery rates and recover better in the allocated model. This could be due to a possible preference of face to face treatments, as the step 2 interventions are more remote and perhaps more technical i.e. telephone, or computerised. This could be a barrier to access or recovery for older people, it is reasonable to predict that this will change as the more technology comfortable generation get older, the current picture is something for services to consider, given the under- representation of the older age group accessing treatment.

Employment status (Appendix 3) at first session is slightly higher than the national IAPT rate of 46%, (53) and shows the proportion of those employed remains consistent at nearly half (47.5% average) across the years, even with a much larger number of participants by year 4. Interestingly for those unemployed there is a proportional decrease from 27% year 1 to 18.5% year 4. In terms of this sample, it could have been caused by the small increase of students each year, and also a larger increase of sick or disabled by year 4 in comparison to year 1 from 6.7% to 14%. These increases are probably attributed to the Academic wellbeing programme

as mentioned previously and the increased targeting of patients with long term conditions by the service. However external factors could affect the proportion registered as unemployed such as more obtaining work, or changes in benefits, therefore not officially registered as unemployed. Similarly to age, the HSCIC (57) health survey of attitudes to mental illness in England found that prejudice, tolerance and support for community care was associated with socio-economic factors, such as education and employment status, with the lowest incomes, those in most deprived areas, no qualifications, unemployed or retired, all had the least positive attitudes towards mental illness. Again this will also likely be a self-reflective attitude and may impact on accessing or engaging with a talking therapy as treatment, for fear of feeling weak or judged by others.

The information on employment status provided by the participant at the last session shows a similar picture to that of the first session, with some very small differences, where some participants' employment or other status has changed during therapy; the majority have remained the same. One of the key performance indicators that IAPT services are measured on is the number of people that move to employment during therapy, with the assumption that their mental health problems are a barrier to obtaining employment. With an average dosage of 6 or 7 sessions (usually weekly) in this sample depending on service model, this may be too high an expectation, that an unemployed participant will seek and find work in less than 2 months from the commencement of therapy. As there is no national IAPT mandate to obtain specific follow up information regarding employment status of all patients sometime after treatment completion, it is difficult to link recovery or improvement in mental health directly to employment change.

8.2 Treatment dosage

As early as 2000, Lovell and Richards (15) argued that some patients get too much treatment and some too little, because not enough is known about the optimal number of sessions for a type of patient. This study observed a fair consistency between cohorts regarding dosage of treatment. This is to be expected with step 2 interventions as they are prescribed and manualised with a set number of sessions. High intensity has much more flexibility, with NICE guidelines advocating 12-20 sessions depending on disorder. Gyani et al (11) found that "services that offered

higher median rates of low intensity and high intensity treatment sessions had overall higher recovery rates." However they do not say what that median is, and whether it is within the range of NICE recommended dosage, particularly for high intensity. The service has always advocates that therapists follow NICE guidelines but has not been prescriptive regarding minimum dosage of step 3, rather the service has advocated a collaborative patient/therapist decision regarding treatment completion, to allow for idiosyncratic adjustments within the treatment plan. It is an interesting result to note that at step 3 only, the average number of sessions is virtually the same in each cohort model type, with one session more occurring in the progression model across all cohorts. Given the anecdotal concern about the progression model in the South area, it could be assumed that more sessions would be delivered, however this is not the case.

With those that have been stepped up, so initially received a step 2 intervention, not recovered and thus received a step 3 intervention, the total dosage is more than step 3 alone, but only by on average 2 sessions. The largest difference appears to be in the progression model rather than allocated, and particularly in the North where the difference between step 3 only and stepped up total dosage is an average of 3 sessions, and between allocated and progression stepped up dosage an increase of an average 2.5 sessions. Therefore it would appear that either through natural treatment end or by drop out, the average step 3 dosages is less than NICE guidelines. This could support the notion that perhaps some patients need less treatment than guidelines advise. (15) Stepped up results show that patients do not receive a much larger dosage of therapy, but on average it increases by 1-2 sessions. This is still slightly less than the minimum recommended NICE guidelines for step 3 of 12 sessions in CBT, however in terms of dosage a combined step 2 and 3 treatment delivery brings the dosage closer to the recommended guidelines. This study has not calculated the specific difference in numbers regarding increased or decreased dosage of treatment at different steps and within models linked to outcomes to work out if there is an effect of dosage on outcomes. This could be worth further exploration.

An interesting comparison can be made with this study's results to the Chan & Adams (44) study where the mean dosage in step 3 is higher (8.95) but lower for low intensity (3.99) than in this study, and overall they reach a higher recovery rate at

52.6%, with a proportional difference of recovery between steps (50% high, 55.3% low). It is possible that as suggested by Gyani et al (11) the higher dose at step 3 effects better recovery rates than have been achieved in this study. However further on in time nationally the HSCIC report (53) which collates statistical data on all IAPT services reported that the majority of completed treatments had 5 or fewer sessions, with the median being 5. This varies in terms of type of disorder, however what is interesting is that the mean number sessions for each disorder is consistently less than the NICE guidelines recommend.

It is difficult to compare the discharge reasons with the Chan & Adams (44) study as their analysis was based on a much smaller sample (n=100) compared to this study (n= 16,723), however proportionally it does appear that the dropout rate in this study is higher, the did not attend (DNA) policy and therefore the probable criteria distinguishing drop out and the attempts to keep patients engaged appears similar.

Seekles et al (40) found a higher dropout rate at lower intensity level treatment, and Firth et al (38) states that "failure to respond to the initial step may actually discourage patients from further engagement across higher steps." This might be a possibility, and is anecdotally one of the concerns raised by clinicians with respect to the progression model, however interestingly this study's results show a reduction dropout rate incrementally through the years, with the North area has the larger decrease, the South only 1 % difference. This may be indicative of the North having 2 years to embed the progression model and therefore patient flow through the care pathway may be quicker. It would have been useful here to consider waiting times for treatment also as a potential influencing factor.

Regression results of the model type cohorts appear to indicate that in comparison to completed treatment the likelihood of dropped out is relatively higher in the progression model, however as is the likelihood of completed treatment, so all that can be inferred here is that all discharge reasons appear more likely in the progression model rather than it meaning that it is more likely to have drop outs, as there has not been a specific regression on the discharge reason cohort to test the predictability of the model type. Receiving a step 2 intervention in the first instance may not be all patients preference of treatment type, however if the treatment is received quicker than a step 3 intervention, it may be that shorter waiting time for

treatment is the contributing factor to treatment completion. What does not appear to be measured in relevant studies is patient preference for specific treatment type preferred over receiving treatment quickly, and the relationship to outcome. Not enough is known about what characteristics or types of people are suited to different types of intervention, and who is more likely to complete treatment and recover. "This raises the key dilemma in stepped care systems of balancing the efficient distribution and composition of organisational resources across various steps and the importance of access to a choice of effective and comprehensive treatments in the early steps." (38) Arguably there is often an assumption that patient choice predicts treatment completion and therefore more likely recovery; however this is a variable that does not appear to be tested in the literature, in terms of patient choice and its relationship with completion and outcome or lack of choice and relationship with dropout or outcome. Although this study did not specifically look at the treatment type or disorder type and number of sessions, the results do appear consistent with the national IAPT picture. Overall there is a question raised around whether the number of sessions recommended by NICE guidelines is, in practice, suitable for all, the evidence would suggest not.

8.3 Psychological measures analysis

Generally the distribution including the mean and standard deviation of first scores appears normal, i.e. the spread looks similar across all 3 measures used, providing some assumption of validity. There is an increase from years 1 to 2, in the score severity. It is likely that by year 2 referrers and the service alike became more familiar with the referring criteria, and from years 2 onwards the severity of referral moved to a more settled picture.

With both the PHQ9 and GAD7 scorings, the detail of change from years 1 to 2 is similar, apart from a small change at PHQ9 moderate level (2.2% decrease), the moderate to severe scores in both measures across the years remaining at a static level. However the bigger change is a decrease in the volume scoring minimal and mild, and a corresponding increase in volume scoring severe in both measures. There it appears that there is an increase in severity of scores in terms of the patients referred to the service. This may be due to the increase of referrals from a wider source, it may be over time referrers increase their understanding and

confidence in the service as they become more familiar. What is important is that around 50% of those entering treatment score in the moderate to severe range which is not the same as the predicted level initially recommended, where it was expected that IAPT would treat as a majority mild to moderate presentations. (46)

The difference in distribution of entry scores in terms of the older age groups has already been mentioned, more specifically the norm for GAD7 scores appears to be fairly consistent across the age groups, however not for PHQ9 and WSAS, where what is noticeable is a norm of less severe scores on both these measures for those aged over 65 onwards. Whilst this could infer that older people are less depressed, and their problems impact less on their functioning, the HSCIC report (53) shows a comparison of IAPT use and specialist mental health services by age distribution demonstrating that the numbers using more specialist service significantly rises with age. This could of course be affected by cognitive impairment disorders such as dementia, and also physical health needs requiring a holistic care co-ordination.

It is more likely that the lower scores are due to perceptions of older people regarding their symptoms and believing them to be symptoms of physical problems only or the aging process rather than recognising them as part of a mental health condition. As mentioned in section 8.1, older people are more likely to have a negative attitude towards mental illness and therefore this could impact on accessing and engaging in treatment.

8.4 Outcomes

Although the progression model was implemented more robustly in years 3 and 4, it would be expected even within this model that some referrals would enter straight into step 3, (i.e. re-referrals previously discharged at step 2) and this is why in the progression model it is not 100% that enter step 2 first.

The findings indicate that the proportion of participants that recovery is greater, in the progression model than the allocated, there is reduced reliable deterioration with progression consistent with all the variations of cohort design. (Table 7, appendix 31).

One explanation could be that regardless of model, the service would have naturally improved over time, and therefore the apparent effect of the progression model is a

coincidence, a cohort effect. Gyani, Shafran, Layard & Clark (11) looked at routine data within year 1 of 32 IAPT services in an observational cohort design, exploring predictors of variability of outcome. The sample size used is similar to whole dataset in this study; however it is comparing 32 sites over the same time period, rather than one site's data over 4 years. Logistical regression is used to test compliance or deviation from NICE treatments, controlling for scores. Findings showed reliable recovery to be at 40.3% with a range variation of 23.9% to 56.5% (SD = 8%). The Improving Access to Psychological Therapies (IAPT) Executive Summary September 2015 published by www.hscic.gov.uk" showed a national mean of the reliable recovery rate at 43.6% for September 2015. Although there has been some small improvement in 3 years, it is still some way from the target of 50% recovery. Clearly there will be some services that have improved over time and some that have not, or have declined, and this shows that service and therapist increased experience does not necessarily directly link with an improvement of recovery rates. There are a number of other factors that may also influence either way. Whilst this study site's staffing profile in high intensity has been stable, therefore those therapists that were training in the first 2 years have gained more experience by the subsequent years. They have not affected the increased recovery outcome, because the progression model was implemented, with an increased volume through low intensity interventions, where the staffing profile was more changeable and a continued level of trainees. This arguably counters any natural service improvement through increased experience.

8.5 High severity needs high intensity?

One of the questions and perhaps criticisms of the progression model that could be levied, and indeed anecdotally clinicians have raised, is that it does not comply with NICE guidelines directly, where the recommendations for certain disorders such as PTSD, social phobia and severe depression (see stepped care diagram figure 1) are to offer high intensity treatment in the first instance, this is stated as typically delivered by step 3 services. This however, is confusing as NICE commissioning guidance places mild to moderate presentation of these disorders as being treated at a step 2 level, and also state "When commissioning services using the stepped care model, commissioners should ensure that local systems allow for some flexibility in how interventions are provided, with the crucial factors being the patterns of local

need." (9) Therefore it is feasible that this can be open to interpretation and be influenced by local capacity, and demand. However to do so, may be open to criticism that NICE guidelines may not be followed. This service has interpreted the NICE guidelines to mean a delivery of stepped care, and has taken the definition within the literature as a progression stepped care model. What is also perhaps confusing for clinicians and service providers reading the evidence regarding stepped care is where the definition and the practice differ. For instance one RCT study (40) sought to compare the outcomes of patients treated in a stepped care model, and those with care as usual. Despite defining stepped care consistent with that of other literature, i.e. low intensity first, the study does not follow this principle for severe presentations. "Even though there is no clear evidence that patients with more severe symptoms of anxiety or depression do not benefit from low intensity (self-help) interventions we decided that patients with more severe disorders should be referred to more specialised mental health care and /or pharmacotherapy and skip the preceding steps." in addition to this, the stepped care participants were randomised to different parts of the stepped care model, to receive different treatments. Where participants are randomised to specific steps it could be argued that the outcomes simply demonstrate the effectiveness of treatments within that step, rather than the effectiveness of a stepped care system. The decision to allocate severe presentations straight to the most specialised treatment perpetuates the idea that severe must need more intensive therapy. Despite acknowledgement that there is no evidence of the ineffectiveness of low intensity, this is an adverse action to the principle and definition of stepped care described, and alongside the study being underpowered, unfortunately weakens the results.

Whilst the discussion here is not questioning the evidence that informs NICE regarding the efficacy of step 3 therapies for clinical mental disorders, the efficacy has been demonstrated for treatment of depression with low intensity interventions (26) and single strand therapies e.g. behaviour activation. (28) However the application of this in practice appears sporadic. Although generally the literature may acknowledge that the efficacy for low intensity interventions treating specific anxiety disorders is less than that of depression, none the less most of studies evaluating IAPT treatment or comparing low to high, include those with clinical levels of anxiety, and there are still demonstrable benefits, and many patients with anxiety treated at

low intensity reaching recovery. (34,44) Although NICE guidelines suggest low intensity interventions can be used for some anxiety disorders displaying mild to moderate symptoms, it recommends not for PTSD or Social Anxiety. (14) From a pragmatic perspective, if a service stringently follows the NICE guidelines to mean moderate to severe anxiety disorders can only be treated by step 3 interventions, and this is applied to this service where 50% were scoring moderate to severe on both measures, it is reasonable to assume that a good proportion of those would meet the criteria for the disorders recommended to be treated at step 3 only. Firstly this forces a service to provide an allocated, matched care model, secondly it requires a service to have enough capacity to meet that clinical demand, which would be at a higher financial cost to employ more high intensity staff, to be able to also meet the IAPT KPI's regarding timescale to enter treatment, and more lately the added waiting time targets. The difficulty for services perhaps is the volume of patients that are referred, and the waiting times for treatment that then occur. To implement this without an increase of resources would simply create a very long waiting list, leaving patients suffering without support, intervention and potentially at risk.

Without good evidence of effective low intensity interventions for PTSD or Social Anxiety disorders, it is not in question that high intensity should be offered, however in routine practice it is highly likely that this will incur a lengthy wait, which could be distressing and exacerbate symptoms for patients. Whilst proper resource funding is of course a necessity, unfortunately the reality for many services may be the volume of demand outweighs the funded capacity. The progression stepped care model used properly could ensure that a larger number of other anxiety disorders are treated with low intensity, and reserving high intensity for PTSD and Social Anxiety. Equally a more generic anxiety low intensity treatment could be offered in the first instance to reduce general anxiety symptoms whilst waiting for a high intensity treatment, under the premise that something is better than nothing. Given the key areas of research need to explore what works for whom, developing a low intensity level treatment for those two disorders may also be a helpful development.

In reality prior to IAPT, the notion that severity means higher more intensive dose of therapy appears to have contributed to long waiting lists, and certainly within this service an allocated model of stepped care has also seen longer waiting lists, which defeats the purpose of IAPT – improving access. Pragmatically a progression model

offers patients a step 2 psycho-educative intervention as the **beginning** of treatment to increase their understanding of their symptoms, and offer practical strategies to reduce symptoms. If this treatment is able to be offered more quickly than a high intensity treatment, and those that do not hit recovery at the end of the step 2 intervention are then stepped up, surely this is a better alternative to patients with complexity sitting on a waiting list with no intervention, and no risk management? NICE guidelines do not purport to keeping patients on lengthy waiting lists without treatment.

Gyani et al (11) showed that compliance with NICE recommended treatments was associated with higher recovery rates, this included low intensity interventions. They found that the factors predicting reliable recovery were initial lower severity "patient's initial PHQ9 and GAD 7 scores had a significant effect on reliable recovery" with higher scores less likely to reach reliable recovery, however higher scores equalled greater reliable improvement. Similarly this study showed that those that score moderate to severe are less likely to recover than lower initial scores.

What is interesting about this study's results showing that overall the progression model achieves an increased recovery rate, this is with a greater proportion receiving only lower intensity interventions than in an allocated model and with an 18.3 % difference between years 2 and 4 of those receiving a low intensity intervention at first session. The results for the moderate to severe PHQ9 initial scores where recovery is higher in the progression model, not only supports the current NICE guidelines for stepped care, but goes further to challenge the notion that severity should be offered a higher intensity interventions first. These results appear to indicate that in reality even those with higher PHQ9 scores will do well with low intensity interventions in the first instance, thus supporting the pure stepped care definition and principle of least intrusive intervention first.

Severe GAD 7 scores also do well in a progression model, and therefore with most receiving a low intensity intervention first also supports the challenge that severity should mean higher intensity. That they are less likely to reliably improve in a progression model than in an allocated model could conversely support that higher intensity should be the treatment offered, as it may be that those that recovered did so once they were given high intensity. Certainly Gyani et al's (11) findings suggest

that more recovery is achieved through services who have a better step up rate from low to high intensity. In this study there is a large proportional difference between years 2 and 4 given a low intensity intervention first. The way this service care pathway is designed and information collated on the database means that it is a single treatment episode that would contain interventions at step 2 and 3, and until 2015 there was no function on the database to be able to easily identify step up. This meant that this study was unable, with such a large dataset to create a variable for those that have been stepped up, and therefore unable to measure this potential effect.

It could be that an anomaly of this particular service, the therapists are more skilled in treating depression, indeed the majority had received training in Behaviour Activation surplus to their low and high intensity courses and this could be contributing to the better results with PHQ9 initial scores. However it does not explain the model difference. It also does not explain the results showing that severe GAD7 scores were more likely to recover in a progression model than allocated. What it does highlight is a further research need to better understand the effective components of treatment for anxiety disorders, potentially at a low intensity level.

Although arguably other unmeasured variables may have contributed to this result, it is an interesting result that questions the idea that complexity and severity require larger doses of therapy than an apparent more simple presentation. These results may pose an alternative perspective that a large proportion of those scoring moderate to severe will do well with simple structured treatments. Perhaps those with extremely complex history and presentation actually may benefit from the simplicity, the uncomplicated nature of a step 2 psycho-educative treatment that guides to do, rather than risk further rumination?

Interestingly Vaillancourt et al (58) found that initial scores are not a factor influencing whether a service achieves a low or high recovery rate, which indicates that there are other factors that do have an influence. Similarly Firth et al (38) systematic review found that “patient severity and symptom chronicity varied considerably, and there were no clear trends that related chronicity/severity to clinical outcome. This is further supported in this study with the outcome of the progression model appearing to achieve better recovery and less deterioration than within an allocated model, in

that implementing a system that removes the allocation of type of treatment based on apparent severity, challenges the notion that severe presentations need higher intensity therapy. As discussed in section 4.2, a study Chan et al (44) found no difference between low or high intensity treatment groups in the baseline psychological measures scores, the high group received more sessions, yet the low intensity group achieved a better recovery rate. It is suggested (44) that it is possible that clinical assessment information finds presentations are different between the groups, with a more complex presentation being allocated a high intensity intervention. This could be a similar occurrence in this study, where although the progression model means low intensity is offered in the first instance; it could be that those that are stepped up are different in clinical presentation. This is not measured in either study, and would be worthy of further exploration, to aid understanding of what works best for whom.

8.6 Allocated versus progression model of stepped care

The major difference between an allocated model and a progression model of stepped care is clinical perspective, principle and practicality.

The allocated model depends on clinician's opinion as to which treatment and level best matches the patient's presentation, based on a belief that the more complex and severe the presentation, the more intensive and complex the therapy needed, delivered by a more qualified and skilled therapist. The clinician's opinion and decision is also informed by the content of training, which may depend on the bias of curriculum creators. Research appealing to the clinician's interest, and as discussed previously, NICE guidelines are all other elements which may support and perpetuate the notion that severity needs high intensity. As also previously discussed, service design provides the framework that also informs and supports the clinician's choice. Service design, in particular IAPT is informed by national IAPT policy and NICE guidelines which describe a mixed model stepped care system, which although advocating least intrusive intervention first, also supports the notion severity means high intensity. The limited research regarding the efficacy of stepped care contributes to the problem of an unclear implementation of stepped care in practice.

Despite apparent agreement in the literature regarding the basic definition of stepped care ([18, 32, 33, 38, 39]), the systematic reviews clearly demonstrate a disparity and broadening of definition which contribute to the heterogeneity regarding a wide range of different models of 'stepped' care, alongside national guidelines (NICE) which appear to advocate a mixed model, that arguably must impact on interpretation and implication of model in routine practice. Inevitably a continuation custom in routine practice, open interpretation of guidelines and a limited body of evidenced based literature, has possibly led to the IAPT programme, a reform of service delivery, continuing to differ across the country within individual services regarding the delivery of stepped care systems. It is likely that there are a good number of IAPT sites that have a mixed model, which will include a proportion of allocated/matched care. "A major problem with this model at present is our lack of clear prognostic determinants with which to match patients to the available treatments."[39] Stepped care has been purported to be an alternative to matched care [33] and recommended as national guidelines and according to van Straten et al [39] "underpins the organisational structure" of the IAPT model design. There is an "assumption of stepped care is that for most patients the low- intensity treatment will be sufficient and only a few will need a higher intensity treatment, thereby making better use of scarce and expensive resources such as therapist time." [39] However the problem with this is the wide interpretation of delivery in routine practice, and indeed in research, as demonstrated by a number of studies ([32,38,39]).

The structure of a progression model is that it removes the clinician choice of intensity of treatment, and requires that all patients will receive a low intensity treatment in the first instance, only being stepped up to a higher intensity if needed. Patient choice of treatment is retained through a range of low intensity treatments on offer included a brief face to face interventions. Whilst method of delivery might differ, i.e. computerised, telephone, psycho education group or brief face to face, the common themes amongst all in this particular service is that they are manualised and standardised treatments, with an emphasis on psycho-education, and motivating the patient to apply what they have learnt. The standardised format means there is less room for treatment 'model drift' which may be a risk with high intensity CBT where the complex patient may be presenting with an apparent number of disorders, that risks the therapist becoming confused as to which is the best model of CBT to

treat with. Clearly the evidence of the efficacy of low intensity interventions for depression and the effectiveness of behaviour activation demonstrates that simple is effective even with severe presentations.

The study (32) that compares 4 different sites delivering their own interpretation of stepped care, not only showed wide variation of implementation, but interestingly the rate of step up from low to high was less than 10% across all sites, which infers that there was perhaps not a demonstrable need for those completing at low intensity to be stepped up, supporting the efficacy of such interventions, and raising the question again that some patients may not actually need high intensity.

This study demonstrates that within one service, greater proportions of participants recover when receiving treatment within a progression model rather than allocated. Gyani et al (11) suggest that better use of stepped care produce better outcomes, and suggested that IAPT services could improve their recovery rates by improving the step up rate. In their review of a number of IAPT services, they found that one of the service characteristics that appeared to predict higher recovery rates were "higher step up rates among individuals who started with low intensity treatments." What isn't clear, is which part of the stepped treatment may have contributed the most to the recovery, i.e. although the reason for step up will be not recovered at the end of a step 2 intervention, what is unknown is whether those participants would have equally recovered with a step 3 intervention alone. The inference that can be made from the observations in this current study is that greater proportions recover in a progression model, with the larger proportion receiving step 2 interventions only. The progression model advocates step up based on need. If the majority of patients in a service recover with a shorter number of sessions, at a lower intensity level, the throughput volume will be greater at the step 2 level, and smaller at the step 3 level, thus likely less waiting times. "Stepped care is a model that seeks to ameliorate problems with access through better allocation of scarce psychological therapy resources". (39) Given the volume of step 2 low intensity treatments within a progression model, the results could be simply evidencing the efficacy of low intensity. The study did not control for this possibility, and could have tested the outcomes for intervention type. However given the service development history and the issues regarding heterogeneous stepped care definition and delivery, the service would perceive the progression model implementation as a necessary mechanism of

change in order to increase the effective use of low intensity interventions, which has then been demonstrated to effect recovery.

The results in this study show that drop out proportion was found to be the same for low intensity treatment only, comparing model type, and at step 3 only and those stepped up were proportionally less in the progression model. There are two possibilities that may explain this occurrence. Either this infers that the progression model may be a more effective service delivery design, or simply a natural service improvement occurred over time. The key component with either possibility could be efficiency improvement, simply getting a treatment to a patient quicker may improve acceptability, and reduce dropout rate. Treating more patients quicker could be achieved either through the system design where treatment more patients with a shorter intervention increases the volume of throughput, or performance management strategies enable therapists to manage their caseload. However as discussed in section 5.4 increasing performance management strategies were in place across the whole service from year 2. Yet there was a difference in outcome between the two hubs. The change that was in place and subsequently the greatest increase in recovery rate occurred afterwards, was when hub B adopted the progression model, and thus the whole service was operating with this delivery design from the start of year 4. So it would appear that the efficiencies made in throughput through the progression model implementation could be linked with the improved dropout rate. Unfortunately this study did not test the significance of dropout rate, no did it control for acceptability of low or high intensity, or waiting times, therefore conclusions still need to remain cautiously optimistic regarding the progression model of stepped care.

8.7 Confounding factors

Certainly across the development of IAPT the issues would have been similar for most services, small numbers of already qualified therapists, large initial recruitment, large numbers of trainees, and more lately movement and attrition of qualified staff. Within this study site in particular these are some of the factors that may have impacted on service improvement and potentially outcomes. As described in section 4.4 Performance management strategies may also have contributed to the trend of increasing recovery. The performance monitoring and management were more

refined from years 2 onwards, with strategies developed to help staff understand the KPI's, i.e. why the service was collecting certain information, these strategies included the distribution of sharing service monthly performance results across teams, and also scorecard to staff with their individual performance, in order to develop a culture of shared understanding and working proactively with performance management tools.

Acknowledging this may have affected outcomes, it is interesting to note is that the gains particularly are made in year 4. It is possible that they took some time for effect to be demonstrated. As performance management strategies were variables not specifically measured in this study, it is perhaps not clear whether the improved use of the progression model or performance management affect the increase in recovery outcome, or a combination of both. However it is important to note that the outcomes remained largely similar for the first 3 years (Table 7), there are very small gains between years 2 and 3, despite increasing performance management strategies. However there was significant increase in recovery in year 4, when the whole service was compliant in delivering the progression model. This suggests that the implementation of the progression model was the most influencing factor of the improved recovery rates.

The notion that trainee therapists becoming qualified and then more skilled and experienced over time effecting recovery is plausible. Gyani et al's (11) study evaluating first year IAPT sites found that where the greater proportion of sessions delivered by staff on pay band 7 or above patients were more likely to reliable recover, this may be less to do with the higher intensity therapy and more to do with experienced therapists, given that the study period was the first year of IAPT and therefore subject to new systems, and a some with a greater proportion of trainees paid on band 6 or below. Therefore where services had a good number of already qualified experienced staff may have skewed the results.

In year 1 of this service delivery, two thirds of the workforce was in training, with one third already qualified as therapists at least one year prior to this local IAPT commencement, and a small number for longer. This could have contributed to the small gains between years 2&3, where new therapists were qualified at end of year 1, and had a further year of close supervision refining skills. It could be expected that

larger gains could have been made at this point, however the service experienced a staff attrition rate from year 2 of between 5-10% each year. This attrition was largely experienced PWP's moving on to high intensity trainee places, or the clinical doctorate in psychology. A service decision as discussed in section 5.3 was taken to replace High Intensity leavers with PWP posts, and a difficulty in recruiting qualified PWPs meant that from year 2 onwards the service continued with around 1/3 of the workforce as trainees, thus a reduction in caseload capacity and throughput, and assumingly smaller recovery rates given the trainee status. This could be a negative effect factor, however despite this possibility, the attrition rate and therefore the recruitment of trainees remained consistent in year 4, however recovery rates rose considerably. "An apparently strong relationship between variables could stem from many sources, including the influence of other, currently unmeasured variables." (59) This study did not collate information regarding the qualification, skill or competence level of therapists delivering the treatment. "Whether an IV appears particularly important in a solution depends on the other IVs in the set." (59) It is possible that the apparent achievement of the progression model with better outcomes in the latter years could be incidental to model type, and may have been influenced by the natural increase of service performance as a phenomenon of natural organisational maturity, or the unmeasured variable of increased therapist skill and competence, or an interaction between all. Indeed a multilevel modelling analysis found that "Therapist effects accounted for 6-7% of outcome variance that was moderated by greater initial symptoms severity, treatment duration, and non-completion of treatment" (38) and also interestingly that "Clinically effective PWPs achieved almost double the change per treatment session", thus supporting both the notion that experienced, competent low intensity therapists, and brief interventions are effective. Furthermore, a study (60) investigating the relationship between CBT competence and patient outcome in routine practice within the IAPT "found little support of a general association between CBT competence and patient outcome; however significantly more patients of the most competent therapists demonstrated a reliable improvement in their symptoms of anxiety than would be expected by chance alone, and fewer experienced no reliable change. Conversely, significantly more patients treated by the least competent therapists experienced a reliable deterioration in their symptoms than would be expected."

Given these findings it would suggest that it is possible that a service with more experienced competent PWPs may positively influence the recovery rate, and a higher rate of least competent therapists may increase the reliable deterioration rate. Consideration does have to be given to the situation in this study where the workforce has always had a proportion of trainees, and new staff due to attrition rates. The increase in recovery rate target nationally in 2013 (see section 5.1) raised the bar for this service at a point it was still refining systems and processes, and therefore the continuation of a proportion of workers with assumed less competence, and a service in continued development may have an effect on the outcomes in either direction.

A fundamental question regarding the results may be are the results achieved through model delivery, or is it evidence of the efficacy of low intensity interventions?

Clearly as discussed previously a stepped care model in name, and following current guidance is more likely to result in an allocated model of care or a mixed model. A purely implemented stepped care, or progression model, by default will result in an increased throughput of patients receiving step 2 interventions. Van Straten et al (33) discussed that the throughput for their brief therapy (broadly equivalent to UK step 2 guided self-help) was less than predicted, and discussed the potential reluctance for therapists to deliver brief therapy to patients with complexity such as personality disorders, thus resulting in therapist bias influencing throughput and model delivery. However the study demonstrated that the majority of patients regardless of severity, or type of disorder do not suffer adverse effects being offered brief therapy. Since then, despite the growing evidence for the efficacy of step 2 interventions, given the mixed model indicated in guidance, and the continuation of allocated care serviced model delivery it would appear that there is a continued belief that severity of presentation needs high intensity therapy, both in terms of dosage and complexity.

8.8 A critique of methodology

The whole dataset used in this study is large, which on the one hand is good as it avoids the risk of small sample bias, and “larger sample sizes increase the ‘power’ of statistical tests.”, making it more likely to detect any existing effect. (61) However, an unnecessarily large sample size may risk uncontrolled variables, and may produce differences that are not meaningful. Tests are performed on various cohorts which

although smaller than the whole dataset are still substantial in size, that there is no issue of small sample bias.

Firstly this study is not a randomised controlled trial (RCT) and therefore by that nature is open to bias and the effect of confounding variables, the results occurring through factors that have not been controlled or measured in this study. “It is always possible that variables unconnected with the independent variable may have produced the changes observed in the dependent variable” (61). Within a RCT, the experimental design would need to ensure that all participants have exactly the same experience, the same treatment delivered in each model to a matched number of ‘demographic type’ of patient, with the same number of treatment sessions, etc. No information was collected on the type of treatment delivered, and there is a range and difference within the steps, for instance computerised cognitive behavioural therapy (cCBT) or telephone guided self-help (TGSH) or face to face guided self-help (FGSH) are delivered very differently, however there does not appear to be a difference in outcome whether treatment is delivered face to face or over the telephone. (29) Step 3 therapies which are not all CBT informed, IPT and EMDR are different modes of therapy to CBT and are only recommended for specific disorders, severe depression for IPT and PTSD for EMDR. (Figure 1). However as mentioned in section 2 in terms of outcomes for depression, all therapies are just about equal. (12,13). Other than the descriptive statistics no further analysis was undertaken on the type of problem being treated, it is entirely possible that this could be a confounding variable to the results. NICE guidelines which are evidence based recommend specific treatments at different levels and dosage, for certain disorders, including level of complexity and chronicity. This information was not collated nor measured, Comparing types of treatment was not the question concerned in this study, as within a stepped care model it is expected there will be a range of treatments, and there is already extensive research regarding the comparison of different therapy types, with ultimately a meta-analysis of studies comparing therapy for depression showing no therapy is particularly more superior than another. (12) Therefore for the basis of the question explored it was not felt necessary to control for treatment type.

Further patient characteristics as well as disorder that could impact on outcomes could be personality traits, history of mental health problems, previous treatments and response, personal factors that impact on mental health such as housing,

relationship issues, family or other support etc. Although the service does collect information about previous mental health difficulties and treatments and response, and within the content of their treatment and subsequent clinical records there would be information of other potential affecting factors, it is not collated or reported on in the same manner as the scores and outcomes, and demographic information. Therefore any future analysis of these factors would require a systematic information collation, either set up through specific research studies exploring specific areas, or through further questionnaires' routinely administered with patients.

Whilst RCT's may be deemed as the 'gold standard' in terms of evidenced research, primarily due to their ability to control variables and demonstrate causation, there may be difficulties in translating results and implications into routine practice, thereby conversely the 'artificial nature' of RCT's with " 'demand characteristics' may distort the procedure and where persons studied are dehumanised." (61) Observational methods explore the 'real world', and with this as a design of a study, "the emphasis is on observation as the main procedure for data gathering, a non-experiment in which records are made of relatively unconstrained behaviour as it occurs." (61)

The study was observational in design, looking at data in routine practice. The use of an observational study design with retrospective data meant that it was not subject to common issues with an observational design, such as bias and influence of participants knowing they are being observed, i.e. the Hawthorne effect. (61) There are common risks of such a study design regarding selection and information bias (62), and there were several elements within the dataset or the design which aims to control bias. All participants were treatment completers, using the KPI definition rather than therapist perspective, the extraction of data from the electronic database which has mandatory fields and the service regularly quality audits meant that there was very little missing data, in an extremely large dataset. Outcomes were derived from calculated scores meeting a certain point i.e. below caseness for recovery, and therefore not withstanding the potential issues regarding reliability of self-reported scores, this information should be perceived as accurate. There could be a possibility of selection bias, regarding defining the cohort groups in the first cohort because there was no exact date where the service completely switched from the allocated to progression model. As described in the methodology rationale, although communication to the workforce regarding model changeover began at 18 months,

consideration of patients who were in the middle of treatment at this point, meant that a reasonable timescale for model adjustment would be at the end of year 2. Awareness of one of the hubs in the service continuing to largely operate the allocate model in year 3, provided an opportunity for this study to have other cohorts to compare, and the sensitivity analysis isolating years 2 and 4 attempts to control for the variable of systems and procedures being embedded in the service during year 1, alongside a greater number of trainees, and impurity of model during year 3, as the South hub continued to operate with the allocated model. Therefore despite the methodological limitations of an observational study design, it is the method in terms of exploring the retrospective data, and based in routine practice. A simple analysis of only the whole service data could have been undertaken, however recognition of the potential variable effect and a need to attempt to lever some control over the confounding variables led to the geographical cohorts, and then further refining the sample size with the sensitivity analysis of isolating years 2 and 4.

As mentioned briefly earlier, sample size of whole dataset is large, and still considerable even with the cleaner, smaller sensitivity sample. This leads to a potentially good effect size, however the statistical tests used are not without their limitations with a large sample.

Parametric techniques assume that the distribution of scores in the population is normal, (51) and there is some argument that even if all the assumptions are not met, with a large sample size they are fairly robust and will tolerate minor violations. This was the case with the chi square results for some areas, and although given the argument about robustness, some data adjustments were made to protect the risk of skew, for instance not undertaking parametric tests on the ethnicity variable, and removing the age outliers for regression tests.

Although arguably the ethnicity spread is representative of the local population, the values of any category other than White British were so small in comparison that any outcome analysis for those categories could not ethically be generalizable.

Similarly the values of the outliers in the age variable may well have been so small that it was tolerable by the regression tests, however the decision was made to remove the outliers, to avoid potential skew of the results.

Normality of distribution would normally be observed through scatterplots, (51, 61) however the sample size was so large visually this was impossible. However through the use of boxplots the range and the norm of each factor within the variable could be observed.

The data set met the criteria for the Chi square tests as each value was unique for each test, and all values were frequencies, i.e. the number of participants recovered. There were some areas where there were low expected frequencies, such as in the age variable, and the rule of Cochran (1954) identifies those where there is more of the 20% of the expected frequency falling below 5. (61) However there is some debate in the statistical literature that 2x2 chi square tests are accurate as long as the total sample size is greater than 20. (61) It is on this basis given the dataset in this study is extremely large that the Chi square tests were accepted, even where the assumptions are not met.

One of the research questions was to explore the relationship baseline factors may have with the outcome. Logistical regression analysis is essentially a set of techniques useful to predict or explain if and by how much a set of independent variables might affect an outcome, a dependent variable. The specific regression technique used will depend on the nature of the outcome variable, the impact of the independent variables is explained through odds ratios. (63)

In this case the outcomes used are reliable recovery, reliable improvement, no change and reliable deterioration. These are categorical variables, and as there is more than two, multinomial regression was considered however initial model fit tests indicated a possible poor fit. Treating the dependant variable, the outcomes as ordinal could mean using ordinal regression. Initial assumption tests showed that the data was a good fit, however once the independent variables were added, in particular the age and employment categories, the dataset failed to meet the proportional odds assumption Therefore it was viewed that the most robust method of regression with this data was to categorise the dependant variable into dichotomous, and run simple binary logistical regression.

Unlike ordinary linear regression, logistical regression does not require normally distributed variables, does not assume linearity of relationship between the independent and dependent variables, or homoscedasticity. However the

independent variables have to have a linear relationship to the logit of the dependent. (63) Outliers may influence logistic regression results (51), in this case, the first and last age categories were removed due to extremely low numbers in both.

The Hosmer-Lemeshow goodness of fit test was undertaken to check that the data did not conflict with assumptions made by the regression model. A large *p* value is viewed as indicative of good model fit. However one of the problems with this test is that with a small sample size, the test gives a high p value as it may lack power to detect mis-specification problems rather than it showing a good fit. This test is preferred compared to the classification tables regarding the assessment of model fit, and it is considered more robust than the chi-square test particularly if the sample size is small. This is not the case with this dataset. There are issues with the Hosmer-Lemeshow test and large data sets however, Kramer & Zimmerman (64) undertook replications of the test 1000 times with a number of studies with large datasets and found that the larger the sample size the more the more times significance was found. This was not the case with this data set, where the majority of the regressions showed the data was a good fit for the model, apart from four sets of regressions that did not meet the Hosmer-Lemeshow test which were the allocated model and reliable improvement outcome, progression model and no change outcome and the initial score group PHQ9 moderate/severe with reliable improvement, and reliable deterioration outcomes. This does not however invalidate results, as there are issues of reliability with the Hosmer–Lemeshow test and unfortunately SPSS does not give another option, but other information needs to be considered alongside. The sizes of sample, the probabilities, both observed and predicted, and adjunct measures of model calibration, for instance are other factors to consider with the reliability of regression model fit. (64)

8.9 Limitations

Although the dataset is large, with missing data at miniscule levels that do not effect results, it is routine practice data, and therefore variables are not as controlled as in a RCT. An RCT within routine practice is viewed as a more reliable method of study, therefore the findings of an observational study with retrospective data should be considered limited. The methodology of the study has attempted to control variables and outliers, for example ethnicity as a variable was excluded from logistic

regression as the numbers in categories other than White British were not large enough for the results to say anything reliable regarding the relationship between this baseline factor and outcomes.

There are limitations with the purity of the cohort design. As described earlier, the move from allocated stepped care to a progression model of stepped care did not happen with a clean shift from one model to the other, rather it was a gradual change impacted by level of adherence to policy and procedure, and culture change issues. It could be argued that the as the dividing lines between the allocated and progression model, were in reality more blurred than a clear fixed point of absolute change, that the model is not a dependent variable. The testing of different cohorts was an attempt to control the impact of those environmental variables, and to test the reliability of the results of the first cohort design. Further sensitivity analysis tests attempted to also provide comparative measures of validity and reliability the first cohort results, and used the model type as an independent variable to test the predictability. However even with all different cohort designs and sensitivity analysis corroborating or improving on the first results, they all are the product of routine practice data, tested with an observational study design, and as such are not as clean as what might be expected within an RCT, and therefore should be viewed with those limitations.

The results provide some interesting implications for psychological therapies service delivery planning, and in particular IAPT services. In terms of cost effectiveness this is likely to provide a strong economic rationale for services with limited resources. If the outcomes are the same or better using a progression model, the model design requires more therapists qualified and able to deliver low intensity interventions compared to high, and the current staffing model within IAPT service is that low intensity interventions are delivered by lower grade (and therefore salary) than high intensity therapists interventions, there is a compelling argument from a simple and crude economic perspective for services to reconfigure to a progression model. However cost effectiveness would need to be analysed taking other factors into account such as treatment dosage, outcomes and referral, and follow up analysis in an RCT.

There are several factors to consider that may have contributed to the result, not least cohort effect; the service over time may simply have got better at delivering brief interventions, therapists more experienced over time may have influenced better recovery. During year 1 of service delivery two thirds of the workforce were in training either as PWPs or high intensity. It could be accepted that during year 2, many were still not finished their training and many were consolidating their skills as newly qualified therapists. What wasn't measured in this study was qualification status and length of experience of therapist as a potential variable. However this may be an impacting variable if indeed the service therapist experience on the whole did rise incrementally. In actual reality whilst that may be the case for the high intensity staff where since year 1, there has not been a huge attrition rate at this level, with high intensity staff remaining within the service. The step 2 workforce however has seen a regular attrition rate every year since year 1, where around 10% of this workforce has left their posts each year, a very small number have been retained within the service as senior PWPs or trainee high intensities, however the majority have actually left the service to obtain senior or trainee posts elsewhere or to undertake study on the psychology clinical doctorate course.

The impact of such as turnover has meant that the service has every year had a larger number of trainee PWPs to replace qualified staff, reducing capacity and presumed skill level. Although these factors have not been measured for the purpose of this study, it is likely that incremental experience of staff does not contribute greatly to the improved recovery rate in the latter years, as this is likely offset by the comparable inexperience and reduced capacity with the regular level of trainee staff.

Environmental factors may have contributed, for instance IAPT as a concept, as a service was new, requiring in some cases quite a radical shift from how mental health therapy was delivered previously, and indeed service design of IAPT is different from previous primary care or mental health teams. Even collecting measures every session was different. As staff adjusted to a new way of working, in a new service, with new systems, it is possible that confidence and competence may be affected by adjustment, and that by the 3rd and 4th year, the culture of the service, and the individual staff were more confident and this impacts positively on competence. This was not measured in this study, and arguably may be a limitation with the findings.

The findings regarding score severity are interesting, and challenge the notion that the higher the score should mean a higher intensity treatment. However a further limitation to this aspect of the study was that clinically assessed information was not included to define severity and it may be that those with higher scores that did not recover, or were stepped up had different clinically assessed information, than those that did recover.

Propensity scoring method was originally a considered method in this study. Propensity matched scoring is often used to control the bias potentially found in observational studies, compared to in a large RCT randomisation will on average balance any bias. However as this method only estimates an average treatment effect and can only account for variables that are observed, in this particular study it would not account for the issues mentioned around environmental factors of cultural change, therapist competence etc. therefore would not necessarily have produced more reliable results than the tests actually used.

8.10 Implications for routine practice

The findings in this study are similar to the national picture (11) where although it seems very obvious, completing treatment or being employed gives a patient a better chance of recovery. Being unemployed, sick or disabled (but not registered disabled) means less of a chance of recovery, and whilst mental health services cannot necessarily influence job provision, closer work with employment coaches, services or people with awareness and links to training opportunities and organisations offering voluntary work could help to tackle those other factors that are clearly linked with mental ill health and impact on recovery. Similarly pathways of joint working or even collaborative care regarding the dovetail between physical health problems and mental health problems may be helpful. Within this service a specialised LTC project operates and achieves reasonable recovery rates.

There is an under- representation of ethnic minorities, and older people accessing this service which is consistent with the national picture, and also to a certain extent younger men. These are areas which are worthy of exploration regarding promotion and engagement to try and increase access.

Given the results that are demonstrated, regarding an increase of 4-10% in recovery with a progression model, and that recovery is more likely in a progression model, it certainly is worthy of consideration for any IAPT service currently delivering a mixed or allocated stepped care model. Given the emphasis for any IAPT service to balance the apparent competing demands of improving access, whilst achieving a minimal target of 50% recovery rate, is clearly a challenge given the average national rate is 43%. There are a number of factors that will impact on this, some of which will be specific to local services, such as staff numbers, staff attrition rates, delivering in higher levels of deprivation which is linked with higher mental health problems, to name a few. However if changing from an allocated model to a progressive model of stepped care can effect an improvement on volume of patients treated, as well at the very least not negatively impacting on recovery rates, then this may be worth a consideration for services seeking to make improvements. This study appears to indicate that for those stepped up in a progression model compared to step 3 only treatment there would be an increase of 1-2 sessions. However if the proportion that receive treatment at step 3 is reduced, through the progression model increasing the volume treated at step 2, it is likely that the actual cost is reduced. A clinical and cost benefits analysis would be useful here to measure the extent of the cost and effect of this.

A further implication for routine practice is the finding that severity does not do worse in a progressive model of stepped care; conversely moderate to severe groups for both psychological measures are more likely to recover in a progression model. This does challenge what appears to be a clinician bias towards the notion that severity should be treated with high intensity. This study can support the growing evidence of the efficacy of low intensity interventions, and also add to the research regarding the model of stepped care. Crucially, because this study is set in routine practice, the findings are of particular value and relevance to IAPT services. Confounding factors are controlled in RCT's, specifically to demonstrate causality; in particular participants are carefully selected according to strict criteria to ensure that the results that are achieved are not affected by any other factor than the one of interest. As this study is observational using retrospective data, its results are potentially more realistic and more applicable to routine practice, where there are naturally issues of variance regarding both patient characteristics i.e.co-morbidity, and service based

variation in terms of competence and experience of clinician. Therefore the results can be seen as realistically demonstrating a relationship between model design and outcome, with variance that within routine practice is expected.

8.11 Conclusion and future research recommendations

This study aimed to review the literature regarding the efficacy of stepped care and analyse observed outcomes of two different models of stepped care, allocated/matched care and pure/progression stepped care.

The literature review demonstrates heterogeneity in definition of stepped care and inclusion of delivery models, which whilst providing some interesting and challenging discussion, potentially hinders rather than helps demonstrate the efficacy of a stepped care model. Given the dearth of studies comparing different models of stepped care, i.e. matched or allocated versus progression, there is a possibility that the studies demonstrating the efficacy of the stepped care model may be more likely demonstrating the efficacy of the treatments delivered within the stepped care model, rather than the effect of the model design itself. There is an assumption made that all IAPTs implement in practice the literature definition of pure stepped care, however there is no evidence of this and that it is likely that nationally IAPTs have a wide range of service model designs, of allocated/matched care, progressive stepped care or a mixture. There is clearly a need in terms of comparing the performance of IAPTs around the country, to do so in the context of understanding each services model design. The national IAPT programme can provide a further research opportunity of numerous services with some of the same components i.e. low and high intensity interventions, to survey each service interpretation of stepped care delivery, map out each service workforce, in terms of qualification and experience of therapists delivering treatment and explore in conjunction with the already collated data regarding outcomes.

Given the limitations discussed, future research would be recommended to build on these findings, with staff experience, competence and confidence measured and variable impact controlled. Controlling environmental factors could be useful, short of setting up two completely new services each with a different model so that

experience and confidence can occur over the same length of time, RCT's within routine practice will always have limitations, and always be impacted by outside environmental factors. An RCT that sets up an allocated model, and a progression model in the same service with the environment the same and staff experience evenly distributed, or an RCT based in multiple services that have defined models that are either allocated or progression, that can control factors such as staff confidence, competence, experience, robustness of systems, could further test out the preliminary findings in this study. A cost benefits analysis within such studies would greatly enrich the findings.

In the absence of a fairly radical study design, exploring the impact of some of the above mentioned factors through a mixed methods study, collating information about clinician competence and experience and adherence to service model, and testing whether there is any relationship between these and clinical outcome could prove useful.

The service does need to improve the robustness of diagnostic criteria, and it would be useful to test out the relationship between disorders, type of treatment, step treated, dosage and clinical outcome. A further study using propensity scoring method could be used to match the participants' characteristics and entry scores to control for those variables. Testing the hypotheses that patient choice effects treatment completion and outcome should be a future consideration, particularly in relation to step and intensity of treatment in the context of understanding better the mechanisms needed for efficiency and maximum outcome attainment within a stepped care system. Looking at the relationship of therapist qualification and competence levels and impact on treatment outcome is another important factor worthy of future study.

Higher score severity was found to not be a negative factor in terms of outcome within the progression model. There is a need to explore further added assessment information regarding perceived complexity and severity and any relationship with treatment step and outcome. An RCT that assessed and recorded clinicians recommended treatment and intensity, and then randomised patients of various identified complexity and severity to low intensity only, high intensity only and low first stepped to high, would be extremely useful to explore the relationship between

clinician assessment, treatment type and intensity, actual treatment and intensity and outcomes.

This study has demonstrated that in this particular service increased recovery rates were obtained in association with delivering treatment to an increased number of participants using a progression model of stepped care. This study has also demonstrated that participants with moderate to severe scores do not appear to deteriorate; rather there is a larger proportion that recovers whilst being treated within a progression model. A higher volume of participants treated at step 2, within progression means that in this service there is a higher proportion of therapy delivery by NHS pay bands 4 &5, compared to the allocated model. By simple association of pay band structure and relatively similar dosage to allocated model for those in step 3 or have been stepped up, it could be argued that the progression model is more efficient and cost effective, as well as achieving better clinical outcomes. A health economic analysis would need to be undertaken to explore this further, as is also suggested by van Straten et al (39) regarding the cost effectiveness of stepped care compared to high intensity alone and matched care.

Fundamentally the problem with the existing evidence is both in research trials and evaluation of routine practice, stepped care has a particular definition but then may be implemented in a more stratified matched care manner, either through design or clinician bias, or a mixture. This is also supported by a mixed and undetailed description of stepped care within the NICE guidelines. There does need to be an acceptance in the routine practice world of the robust evidence of low intensity interventions, which form the basis of a pure stepped care model. There also needs to be further clarity in the guidelines to form a synergy between definition and recommended implication.

The systematic review and meta-analysis of stepped care recommended for future research “the ideal test, against true matched care or against high intensity care for all patients has not been performed yet.” (39)

This study cannot claim to be that ideal test, given it is not an RCT, however it has taken the unique opportunity that has naturally occurred within the development of a particular IAPT service, and explores the impact of a change of service delivery model on outcomes adding to a small volume of literature. Not enough is known

about the components of treatment or the optimal dosage that work for different patients to deliver an effective matched care model that is truly without bias, and purely evidenced based to be able to treat the patient's needs. This study demonstrates that there is a significant likelihood of more patients recovered within a progression model of stepped care, compared to one that operated more akin to matched care, and furthermore those scoring moderate to severe on both psychological measures also are significantly more likely to recover in a progression model. Add together this, the existing robust evidence for the efficacy of step 2 interventions for depression, and the demonstration in this study that anxiety presentations also recover well in a progression model; perhaps a pure stepped care progression model could be viewed, the optimal model in terms of the balance between evidenced based effective treatment and efficiency, until further research finds the mechanisms of effective treatment with an effective system of matching those to the right patient. The implications for clinical practice are a clearer definition of the stepped care model and how it should be implemented, evidence of improved outcomes within a progression model of stepped care, that can be used to consider, influence and improve service design and policy.

References

(1) Choice CS, Team A. IAPT positive practice guide.

(2) Layard R. The depression report: A new deal for depression and anxiety disorders. Centre for Economic Performance, LSE; 2006 Jun.

(3) Layard R, Clark D, Knapp M, Mayraz G. Cost-benefit analysis of psychological therapy. National Institute Economic Review. 2007 Oct 1;202(1):90-8.

(4) NICE N. Anxiety: management of anxiety (panic disorder, with or without agoraphobia, and generalised anxiety disorder) in adults in primary, secondary and community care. London: National Collaborating Centre for primary care. 2004.

(5) NICE N. Depression: Management of depression in primary and secondary care. Clinical Guideline 23. London: National Institute for Health and Clinical Excellence. 2004.

(6) Gene-Cos N. Post-Traumatic Stress Disorder: The Management of PTSD in Adults and Children in Primary and Secondary Care. The Psychiatrist. 2006 Sep 1;30(9):357-.

(7) National Institute for Clinical Excellence. Computerised cognitive behaviour therapy for depression and anxiety: Technology Appraisal 97. National Institute for Clinical Excellence, London. 2006.

(8) Depression NI. The treatment and management of depression in adults (CG90). National Institute for Health and Clinical Excellence. 2009.

(9) National Collaborating Centre for Mental Health (Great Britain), Royal College of Psychiatrists. Common mental health disorders: Identification and pathways to care. RCPsych Publications; 2011.

(10) National Institute for Clinical Excellence, National Institute for Clinical Excellence. Generalised anxiety disorder and panic disorder (with or without agoraphobia) in adults: Management in primary, secondary and community care. London: National Collaborating Centre for Mental Health. 2011.

(11) Clark DM, Gyani A, Layard R, Shafran R. Enhancing recovery rates: lessons from year one of the English'Improving access to psychological therapies' programme.

(12) Barth J, Munder T, Gerger H, Nüesch E, Trelle S, Znoj H, Jüni P, Cuijpers P. Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. PLoS Med. 2013 May 28;10(5):e1001454.

(13) Cuijpers P. Psychotherapies for adult depression: recent developments. Current opinion in psychiatry. 2015 Jan 1;28(1):24-9

(14) Clark DM. Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: the IAPT experience. International Review of Psychiatry. 2011 Aug 1;23(4):318-27

(15) Lovell K, Richards D. Multiple access points and levels of entry (MAPLE): ensuring choice, accessibility and equity for CBT services. Behavioural and Cognitive Psychotherapy. 2000 Oct 1;28(04):379-91.

(16) Scogin FR, Hanson A, Welsh D. Self-administered treatment in stepped-care models of depression treatment. Journal of clinical psychology. 2003 Mar 1;59(3):341-9.

(17) Pilling S, Harvey P. NICE recommends a stepped care approach to managing depression. Guidelines in practice. Accessed from http://www.guidelinesinpractice.co.uk/feb_05_pilling_depression_feb05 [Accessed 19th June 2015]

(18) Bower P, Gilbody S. Stepped care in psychological therapies: access, effectiveness and efficiency. The British Journal of Psychiatry. 2005 Jan 1;186(1):11-7.

(19) Department of Health. Improving access to psychological therapies. Implementation plan: national guidelines for regional delivery. London: Department of Health; 2008.

(20) Clark DM, Layard R, Smithies R, Richards DA, Suckling R, Wright B. Improving access to psychological therapy: Initial evaluation of two UK demonstration sites. Behaviour research and therapy. 2009 Nov 30;47(11):910-20

(21) Glover G, Webb M, Evison F. Improving access to psychological therapies: A review of the progress made by sites in the first rollout year. North East Public Health Observatory. 2010 Jul.

(22) Department of Health. Improving Access to Psychological Therapies. The IAPT Data Handbook. London: Department of Health; 2011.

(23) Kroenke K, Spitzer RL, Williams JB. The Phq-9. Journal of general internal medicine. 2001 Sep 1;16(9):606-13.

(24) Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Archives of internal medicine. 2006 May 22;166(10):1092-7.

(25) Davison GC. Stepped care: doing more with less?. Journal of consulting and clinical psychology. 2000 Aug;68(4):580.

(26) Bower P, Kontopantelis E, Sutton A, Kendrick T, Richards DA, Gilbody S, Knowles S, Cuijpers P, Andersson G, Christensen H, Meyer B. Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data.

(27) Ekers D, Richards D, Gilbody S. A meta-analysis of randomized trials of behavioural treatment of depression. Psychological medicine. 2008 May 1;38(05):611-23.

(28) Ekers D, Webster L, Van Straten A, Cuijpers P, Richards D, Gilbody S. Behavioural activation for depression; an update of meta-analysis of effectiveness and sub group analysis. PloS one. 2014 Jun 17;9(6):e100100

(29) Hammond GC, Croudace TJ, Radhakrishnan M, Lafortune L, Watson A, McMillan-Shields F, Jones PB. Comparative effectiveness of cognitive therapies delivered face-to-face or over the telephone: an observational study using propensity methods. PloS one. 2012 Sep 28;7(9):e42916.

(30) Archer J, Bower P, Gilbody S, Lovell K, Richards D, Gask L, Dickens C, Coventry P. Collaborative care for depression and anxiety problems. Cochrane Database Syst Rev. 2012 Oct 17;10(10).

(31) Ekers D, Murphy R, Archer J, Ebenezer C, Kemp D, Gilbody S. Nurse-delivered collaborative care for depression and long-term physical conditions: a systematic review and meta-analysis. Journal of affective disorders. 2013 Jul 31;149(1):14-22.

(32) Richards DA, Bower P, Pagel C, Weaver A, Utley M, Cape J, Pilling S, Lovell K, Gilbody S, Leibowitz J, Owens L. Delivering stepped care: an analysis of implementation in routine practice. Implement Sci. 2012 Jan 16;7(3):5908-7.

(33) Van Straten A, Tiemens B, Hakkaart L, Nolen WA, Donker MC. Stepped care vs. matched care for mood and anxiety disorders: a randomized trial in routine practice. Acta Psychiatrica Scandinavica. 2006 Jun 1;113(6):468-76.

(34) Richards DA, Borglin G. Implementation of psychological therapies for anxiety and depression in routine practice: two year prospective cohort study. Journal of affective disorders. 2011 Sep 30;133(1):51-60.

(35) Higgins JP, Green S, editors. Cochrane handbook for systematic reviews of interventions. Chichester: Wiley-Blackwell; 2008 Feb.

(36) Cartwright N. Are RCTs the gold standard?. BioSocieties. 2007 Mar 1;2(1):11-20.

(37) Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. New England Journal of Medicine. 2000 Jun 22;342(25):1887-92.

(38) Firth N, Barkham M, Kellett S, Saxon D. Therapist effects and moderators of effectiveness and efficiency in psychological wellbeing practitioners: A multilevel modelling analysis. Behaviour research and therapy. 2015 Jun 30;69:54-62.

(39) Van Straten A, Hill J, Richards DA, Cuijpers P. Stepped care treatment delivery for depression: a systematic review and meta-analysis. Psychological medicine. 2015;45(02):231-46.

(40) Seekles W, van Straten A, Beekman A, van Marwijk H, Cuijpers P. Stepped care treatment for depression and anxiety in primary care. a randomized controlled trial. Trials. 2011 Jul 7;12(1):171.

(41) Singh J. Critical appraisal skills programme. Journal of Pharmacology and pharmacotherapeutics. 2013 Jan 1;4(1):76.

(42) Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Annals of internal medicine. 2009 Aug 18;151(4):264-9.

(43) Weich S, Lewis G. Poverty, unemployment, and common mental disorders: population based cohort study. Bmj. 1998 Jul 11;317(7151):115-9.

(44) Chan SW, Adams M. Service Use, Drop-Out Rate and Clinical Outcomes: A Comparison Between High and Low Intensity Treatments in an IAPT Service. Behavioural and cognitive psychotherapy. 2014 Nov 1;42(06):747-59.

(45) Cuijpers P, Hollon SD, van Straten A, Bockting C, Berking M, Andersson G. Does cognitive behaviour therapy have an enduring effect that is superior to keeping patients on continuation pharmacotherapy? A meta-analysis. BMJ open. 2013 Jan 1;3(4):e002542.

(46) Glover G. Estimating the prevalence of common mental health problems in PCTs in England: A first approximation of the expected caseload for new psychological therapy services.

(47) Barnfield TV, Mathieson FM, Beaumont GR. Assessing the development of competence during postgraduate cognitive-behavioral therapy training. Journal of Cognitive Psychotherapy. 2007 Jun 1;21(2):140-7.

(48) Smith PC, Mossialos E, Papanicolas I. Performance measurement for health system improvement: experiences, challenges and prospects.

(49) Propper C, Wilson D. The use and usefulness of performance measures in the public sector. Oxford review of economic policy. 2003 Jun 1;19(2):250-67.

(50) Mundt JC, Marks IM, Shear MK, Greist JM. The Work and Social Adjustment Scale: a simple measure of impairment in functioning. The British Journal of Psychiatry. 2002 May 1;180(5):461-4.

(51) Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). British Journal of Clinical Psychology. 2007 Mar 1;46(1):1-8.

(52) Bradford-ONS S. Ethnicity and National Identity in England and Wales 2011.

(53) Health & Social Care Information Centre. Psychological Therapies, Annual Report on the use of IAPT services: England – 2013/14 Experimental Statistics. 2014.

(54) Department of Health. Preventing Suicide in England: A Cross government outcomes strategy to save lives: Assessment of impacy on equalities. London: Department of Health 2012.

(55) Wylie C, Platt S, Brownie J, Chandler A. Men, suicide and society. London: Samaritans. 2012.

(56) Iapt.nhs.uk, (2016). IAPT | Equalities > Older People. [online] Available at: http://www.iapt.nhs.uk/equalities/older-people [Accessed 9 Feb. 2016].

(57) Ilic N, Henderson C, Henderson C, Evans-Lacko S, Thornicroft G. Attitudes towards mental illness.

(58) Vaillancourt K, Manley J, McNulty N. Why has our recovery rate dropped? An audit examining waiting times, starting scores and length of treatment in relation to recovery within an IAPT service. The Cognitive Behaviour Therapist. 2015;8:e7.

(59) Tabachnick BG, Fidell LS. Multilevel linear modeling. Using multivariate statistics. 2007:781-857.

(60) Branson A, Shafran R, Myles P. Investigating the relationship between competence and patient outcome with CBT. Behaviour research and therapy. 2015 May 31;68:19-26.

(61) Coolican H. Research methods and statistics in psychology. Hodder & Stoughton Educational; 1990.

(62) Jepsen P, Johnsen SP, Gillman MW, Sørensen HT. Interpretation of observational studies. Heart. 2004 Aug 1;90(8):956-60.

(63) Restore.ac.uk, (2016). *4.12 The SPSS Logistic Regression Output*. [online] Available at: http://www.restore.ac.uk/srme/www/fac/soc/wie/research-new/srme/modules/mod4/12/index.html [Accessed 9 Feb. 2016].

(64) Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited*. Critical care medicine. 2007 Sep 1;35(9):2052-6.

Impact of a Progressive Stepped Care Approach in an Improving Access to Psychological Therapies Service: An Observational Study

Volume 2/2

Appendices

Lisa Susan Boyd

Thesis submitted for the degree of Master of Science in Research in The School of Medicine and Health at Durham University.

February 2016

Appendices contents

Appendix 1: Other Database Searches

15/12/14 other databases through HDAS

1. PsycINFO; (depression OR depressive OR affective OR "low mood").ti,ab; 258133 results.

2. PsycINFO; ATYPICAL DEPRESSION/ OR BECK DEPRESSION INVENTORY/ OR "DEPRESSION (EMOTION)"/ OR "LONG-TERM DEPRESSION (NEURONAL)"/ OR MAJOR DEPRESSION/ OR POSTPARTUM DEPRESSION/ OR RECURRENT DEPRESSION/ OR SPREADING DEPRESSION/ OR TREATMENT RESISTANT DEPRESSION/; 116461 results.

3. PsycINFO; 1 OR 2; 266461 results.

4. PsycINFO; ("stepped care" OR "collaborative care" OR "stratified care" OR "matched care").ti,ab; 1153 results.

5. PsycINFO; 3 AND 4; 470 results.

6. PsycINFO; (RCT* OR "randomi* controlled trial*" OR "systematic review*" OR "meta analys*").ti,ab; 41147 results.

7. PsycINFO; 5 AND 6; 133 results.

8. PsycINFO; (observational OR naturalistic OR "quasi experiment*" OR "natural experiment*" OR "before and after study*" OR "service evaluation*").ti,ab; 35451 results.

9. PsycINFO; 5 AND 8; 11 results.

10. PsycINFO; (anxiety OR anxious).ti,ab; 144534 results.

11. PsycINFO; ANXIETY/ OR ANXIETY DISORDERS/ OR ANXIETY MANAGEMENT/; 58540 results.

12. PsycINFO; 10 OR 11; 149939 results.

13. PsycINFO; 4 AND 12; 157 results.

14. PsycINFO; 6 OR 8; 75008 results.

15. PsycINFO; 13 AND 14; 40 results.

16. PsycINFO; 5 AND 14; 142 results.

17. PsycINFO; (("stepped care" OR "stratified care" OR "matched care")).ti,ab; 469 results.

18. PsycINFO; 3 AND 17; 152 results.

19. PsycINFO; 14 AND 18; 59 results.

20. PsycINFO; 12 AND 17; 82 results.

21. PsycINFO; 14 AND 20; 28 results.

22. MEDLINE; (depression OR depressive OR affective OR "low mood").ti,ab; 295460 results.

23. MEDLINE; DEPRESSION/; 82520 results.

24. MEDLINE; 22 OR 23; 318398 results.

25. MEDLINE; ("stepped care" OR "stratified care" OR "matched care").ti,ab; 878 results.

26. MEDLINE; 24 AND 25; 217 results.

27. MEDLINE; (RCT* OR "randomi* controlled trial*" OR "systematic review*" OR "meta analys*" OR observational OR naturalistic OR "quasi experiment*" OR "natural experiment*" OR "before and after study*" OR "service evaluation*").ti,ab; 282400 results.

28. MEDLINE; 26 AND 27; 83 results.

29. MEDLINE; (anxiety OR anxious).ti,ab; 126267 results.

30. MEDLINE; ANXIETY/ OR ANXIETY DISORDERS/; 79809 results.

31. MEDLINE; 29 OR 30; 150854 results.

32. MEDLINE; 25 AND 31; 120 results.

33. MEDLINE; 27 AND 32; 48 results.

34. CINAHL; (depression OR depressive OR affective OR "low mood").ti,ab; 51403 results.

35. CINAHL; DEPRESSION/ OR DEPRESSION, REACTIVE/; 43592 results.

36. CINAHL; 34 OR 35; 66300 results.

37. CINAHL; ("stepped care" OR "stratified care" OR "matched care").ti,ab; 224 results.

38. CINAHL; 36 AND 37; 67 results.

39. CINAHL; (RCT* OR "randomi* controlled trial*" OR "systematic review*" OR "meta analys*" OR observational OR naturalistic OR "quasi experiment*" OR "natural experiment*" OR "before and after study*" OR "service evaluation*").ti,ab; 72364 results.

40. CINAHL; 38 AND 39; 29 results.

41. CINAHL; (anxiety OR anxious).ti,ab; 25986 results.

42. CINAHL; ANXIETY/ OR ANXIETY DISORDERS/; 18694 results.

43. CINAHL; 41 OR 42; 33143 results.

44. CINAHL; 37 AND 43; 27 results.

45. CINAHL; 39 AND 44; 14 results.

46. EMBASE; (depression OR depressive OR affective OR "low mood").ti,ab; 352391 results.

48. EMBASE; ("stepped care" OR "stratified care" OR "matched care").ti,ab; 1039 results.

49. EMBASE; (RCT* OR "randomi* controlled trial*" OR "systematic review*" OR "meta analys*" OR observational OR naturalistic OR "quasi experiment*" OR "natural experiment*" OR "before and after study*" OR "service evaluation*").ti,ab; 346976 results.

50. EMBASE; DEPRESSION/ OR LONG TERM DEPRESSION/ OR MAJOR DEPRESSION/ OR MIXED ANXIETY AND DEPRESSION/; 281144 results.

51. EMBASE; 46 OR 50; 447419 results.

52. EMBASE; 48 AND 51; 291 results.

53. EMBASE; 49 AND 52; 97 results.

54. EMBASE; (anxiety OR anxious).ti,ab; 160979 results.

55. EMBASE; ANXIETY/ OR ANXIETY DISORDER/ OR ANXIETY NEUROSIS/ OR MIXED ANXIETY AND DEPRESSION/; 168811 results.

56. EMBASE; 54 OR 55; 215768 results.

57. EMBASE; ("stepped care" OR "stratified care" OR "matched care").ti,ab; 1039 results.

58. EMBASE; 56 AND 57; 156 results.

59. EMBASE; (RCT* OR "randomi* controlled trial*" OR "systematic review*" OR "meta analys*" OR observational OR naturalistic OR "quasi experiment*" OR "natural experiment*" OR "before and after study*" OR "service evaluation*").ti,ab; 346976 results.

60. EMBASE; 58 AND 59; 62 results.

Appendix 2: Summary of Literature Review Findings

Authors and date	Study design and participants	Findings and limitations
Richards, Bower, Pagel, Weaver, Utley, Cape, Pilling, Lovell, Gilbody, Leibowitz, Owens, Paxton, Hennessy, Simpson,Gallivan, Tomson, Vasilakis 2012	Observational in routine practice n=7698 An observational study using data of patients in 4 different routine health settings operating stepped care models, with an analysis of the proportion of patients accessing treatment, at which 'step' and the transitions between steps. Each service developed their interpretation of stepped care, and data was collected and analysed.	The interpretation of the NICE guidelines for stepped care was implemented with large variation across all 4 sites, with a particular difference with the ratio of low or high intensity treatments received by differing proportions of patients across all the sites. Findings indicate that service model type may have a relationship with the volume of patients accessing high or low intensity treatment. Clinical outcomes data was not present to be able to analyse.
Chan S. W.Y., Adams,M ., 2014	Between groups design, sample analysis n=100 This is a small sample study comparing low and high intensity treatments, analysing outcomes and dropout rates.	There is no difference between groups regarding dropout rate, a level of % difference between groups in terms of recovery rates – 50% high intensity, 55.3% low intensity, however no significant statistical difference. Limitations include potential bias In the randomisation process as the variable of number of contacts was not controlled, and it is possible that those with a higher number of contacts were more likely to be selected.
van straten, A., Hill,J., Richards,D., Cuijpers,P. 2014	Systematic review & meta-analysis of 14 studies (n=5194, 2560 in stepped care) of which 10 are used in a meta –analysis (n =4580, with n= 2243 in stepped care).	7 studies are regarding the delivery of a collaborative care model, 6 studies are regarding an increasing intensity of stepped care with just 2 studies with progressive intensity of stepped care. The review and meta-analysis finds that there is wide heterogeneity amongst design and delivery of service models and stepped care has a moderate effect on depression with the progressive intensity doing significantly worse. Limitations include the wide inclusion criteria that arguably contribute to heterogeneity, and only two studies representing 'true' stepped care, therefore findings should be viewed cautiously.
Firth,N., Barkham,M., kellett,S., 2015	Systematic review n=14 studies. 9 randomised controlled studies, 1 quasi-randomised comparison study, and 3 uncontrolled prospective cohort studies. The number of patients per study range from 18-7859, (mdn = 430).	The review found heterogeneity amongst the stepped care systems. Recovery rates for depression are between 50%-60% in stepped care, and "equivalence to usual care is suggested by comparison studies.", however the evidence in some studies suggesting the superiority of stepped care, the authors find to be inconclusive. Limitations include difficulty in comparison and calculating effect due to wide variation of stepped care systems and study methodology, including two underpowered studies.

Appendix 3: Demographics by year

		Year One		Year Two		Year Three		Year Four		Total	
		N	%	N	%	N	%	N	%	N	%
Gender	Male	702	37.1	1590	37,1	2058	38.2	1867	36.3	6217	37.2
	Female	1191	62.9	2699	62.9	3333	61.8	3278	63.7	10501	62.8
	Not Stated	0	0.0	2	0.0	3	0.1	0	0.0	5	0.0
	Total	1893	11.3	4291	25.7	5394	32.3	5145	30.8	16723	100.0
Age	16-17	0	0.0	0	0.0	0	0.0	3	0.1	3	0.0
	18-24	90	4.8	315	7.3	554	10.3	705	13.7	1664	10.0
	25-34	447	23.6	1019	23.7	1256	23.3	1211	23.5	3933	23.5
	35-44	482	25.5	983	22.9	1220	22.6	1180	22.9	3865	23.1
	45-54	435	23.0	991	23.1	1182	21.9	1086	21.1	3694	22.1
	55-64	297	15.7	689	16.1	866	16.1	670	13.0	2522	15.1
	65-74	116	6.1	242	5.6	249	4.6	218	4.2	825	4.9
	75-84	21	1.1	45	1.0	55	1.0	60	1.2	181	1.1
	85-94	5	0.3	7	0.2	12	0.2	11	0.2	35	0.2
	95+	0	0.0	0	0.0	0	0.0	1	0.0	1	0.0
	Total	1893	11.3	4291	25.7	5394	32.3	5145	30.8	16723	100.0
Ethnicity	British	685	36.2	2564	59.8	4741	87.9	4809	93.5	12799	76.5
	Irish	1	0.1	3	0.1	7	0.1	7	0.1	18	0.1
	Other White background	3	0.2	15	0.3	33	0.6	50	1.0	101	0.6
	White & Black Caribbean	2	0.1	1	0.0	13	0.2	7	0.1	23	0.1
	White & Black African	0	0.0	1	0.0	0	0.0	1	0.0	2	0.0
	White & Asian	0	0.0	2	0.0	5	0.1	1	0.0	8	0.0
	Other mixed background	1	0.1	2	0.0	7	0.1	6	0.1	16	0.1
	Indian	2	0.1	10	0.2	18	0.3	14	0.3	44	0.3
	Pakistani	0	0.0	1	0.0	3	0.1	4	0.1	8	0.0
	Bangladeshi	0	0.0	1	0.0	1	0.0	0	0.0	2	0.0
	Other Asian background	0	0.0	3	0.1	8	0.1	7	0.1	18	0.1
	Caribbean	1	0.1	1	0.0	2	0.0	5	0.1	9	0.1
	African	1	0.1	0	0.0	4	0.1	2	0.0	7	0.0
	Other Black background	3	0.2	1	0.0	3	0.1	0	0.0	7	0.0
	Chinese	0	0.0	9	0.2	21	0.4	11	0.2	14	0.1
	Other ethnic group	4	0.0	3	0.1	7	0.1	15	0.3	56	0.3
	Not Stated	1190	62.9	1674	39.0	521	9.7	206	4.0	3591	20.9
	Total	1893	11.3	4291	25.7	5394	32.3	5145	30.8	16723	100.0
Disability	Yes	61	3.2	185	4.3	536	9.9	494	9.6	1276	7.6
	No	1832	96.8	4106	95.7	4858	90.1	4651	90.4	15447	92.4
	Total	1893	11.3	4291	25.7	5394	32.3	5145	30.8	16723	100.0
Referred Problem	Mental & Behavioural Disorder due to alcohol use	1	0.1	1	0.0	4	0.1	3	0.1	9	0.1
	Bipolar Affective Disorder	0	0.0	2	0.0	6	0.1	8	0.2	16	0.1
	Depressive Episode	437	23.1	939	21.9	1362	25.3	1144	22.2	3882	23.2
	Recurrent Depressive Episode	142	7.5	330	7.7	389	7.2	273	5.3	1134	6.8
	Dysthymia		0.0		0.0	5	0.1	7	0.1	12	0.1
	Agoraphobia	16	0.8	32	0.7	31	0.6	24	0.5	103	0.6
	Social Phobias	24	1.3	21	0.5	64	1.2	71	1.4	180	1.1
	Specific Phobias	26	1.4	33	0.8	65	1.2	49	1.0	173	1.0
	Panic Disorder	38	2.0	195	4.5	293	5.4	229	4.5	755	4.5
	Generalized Anxiety Disorder	522	27.6	1130	26.3	999	18.5	664	12.9	3315	19.8
	Mixed Anxiety & Depressive Disorder	507	26.8	1310	30.5	1625	30.1	2037	39.6	5479	32.8
	Obsessive Compulsive Disorder	33	1.7	104	2.4	182	3.4	143	2.8	462	2.8

	Post Traumatic Stress Disorder	24	1.3	59	1.4	158	2.9	144	2.8	385	2.3
	Adjustment Disorder	1	0.1	3	0.1	35	0.6	40	0.8	79	0.5
	Somatoform Disorder	1	0.1	1	0.0	22	0.4	25	0.5	49	0.3
	Hypochondriac Disorder	3	0.2	3	0.1	3	0.1	3	0.1	12	0.1
	Eating Disorder	5	0.3	26	0.6	34	0.6	24	0.5	89	0.5
	Other Mental Disorder	63	3.3	91	2.1	103	1.9	250	4.9	507	3.0
	Disappearance or Death of Family Member	2	0.1	0	0.0	14	0.3	7	0.1	23	0.1
	Not Stated	48	2.5	11	0.3	0	0.0	0	0.0	59	0.4
	Total	1893	11.3	4291	25.7	5394	32.3	5145	30.8	16723	100.0
Employment Status First	Employed	922	48.7	2011	46.9	2548	47.2	2455	47.7	7936	47.5
	Unemployed & Seeking Work	511	27.0	1054	24.6	1192	22.1	950	18.5	3707	22.2
	Students Not Seeking Work	92	4.9	269	6.3	361	6.7	378	7.3	1100	6.6
	Sick or Disabled	127	6.7	399	9.3	617	11.4	720	14.0	1863	11.1
	Homemaker	90	4.8	258	6.0	287	5.3	267	5.2	902	5.4
	No Benefits, Not Working or Seeking Work		0.0	5	0.1	20	0.4	12	0.2	37	0.2
	Voluntary Work		0.0	1	0.0	3	0.1	3	0.1	7	0.0
	Retired	121	6.4	269	6.3	342	6.3	329	6.4	1061	6.3
	Not Stated	30	1.6	25	0.6	24	0.4	31	0.6	110	0.7
	Total	1893	11.3	4291	25.7	5394	32.3	5145	30.8	16723	100.0

Appendix 4: Demographics by locality

		North				South				Total	
		Allocated		Progressive		Allocated		Progressive			
		N	%	N	%	N	%	N	%	N	%
Gender	Male	1316.0	37.2	2218.0	37.7	1808.0	37.2	818.0	35.4	6160.0	37.1
	Female	2217.0	62.7	3665.0	62.3	3053.0	62.8	1490.0	64.6	10425.0	62.8
	Not Stated	1.0	0.0	2.0	0.0	2.0	0.0	0.0	0.0	5.0	0.0
	Total	3534.0	21.3	5885.0	35.5	4863.0	29.3	2308.0	13.9	16590.0	100.0
Age	16-17	0.0	0.0	2.0	0.0	0.0	0.0	1.0	0.0	3.0	0.0
	18-24	253.0	7.2	756.0	12.8	361.0	7.4	269.0	11.7	1639.0	9.9
	25-34	869.0	24.6	1412.0	24.0	1098.0	2206.0	529.0	22.9	3908.0	23.6
	35-44	831.0	23.5	1321.0	22.4	1132.0	23.3	553.0	24.0	3837.0	23.1
	45-54	794.0	22.5	1227.0	20.8	1130.0	23.2	516.0	22.4	3667.0	22.1
	55-64	536.0	15.2	841.0	14.3	821.0	16.9	308.0	13.3	2506.0	15.1
	65-74	199.0	5.6	234.0	4.0	273.0	5.6	110.0	4.8	816.0	4.9
	75-84	40.0	1.1	77.0	1.3	43.0	0.9	19.0	0.8	179.0	1.1
	85-94	12.0	0.3	14.0	0.2	5.0	0.1	3.0	0.1	34.0	0.2
	95+	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
	Total	3534.0	21.3	5885.0	35.5	4863.0	29.3	2308.0	13.9	16590.0	100.0
Ethnicity	British	1496.0	42.3	5326.0	90.5	3741.0	76.9	2143.0	92.9	12706.0	76.6
	Irish	3.0	0.1	8.0	0.1	2.0	0.0	5.0	0.2	18.0	0.1
	Other White background	7.0	0.2	62.0	1.1	21.0	0.4	11.0	0.5	101.0	0.6
	White & Black Caribbean	2.0	0.1	9.0	0.2	7.0	0.1	5.0	0.2	23.0	0.1
	White & Black African	1.0	0.0	1.0	0.0		0.0		0.0	2.0	0.0
	White & Asian	1.0	0.0	4.0	0.1	3.0	0.1		0.0	8.0	0.0
	Other mixed background	1.0	0.0	12.0	0.2	2.0	0.0	1.0	0.0	16.0	0.1
	Indian	7.0	0.2	16.0	0.3	12.0	0.2	9.0	0.4	44.0	0.3
	Pakistani	0.0	0.0	4.0	0.1	3.0	0.1	1.0	0.0	8.0	0.0
	Bangladeshi	0.0	0.0		0.0	2.0	0.0		0.0	2.0	0.0
	Other Asian background	0.0	0.0	10.0	0.2	6.0	0.1	2.0	0.1	18.0	0.1
	Caribbean	0.0	0.0	4.0	0.1	4.0	0.1	1.0	0.0	9.0	0.1
	African	0.0	0.0	3.0	0.1	2.0	0.0	2.0	0.1	7.0	0.0
	Other Black background	1.0	0.0		0.0	5.0	0.1		0.0	6.0	0.0
	Chinese	4.0	0.1	7.0	0.1	2.0	0.0		0.0	13.0	0.1
	Other ethnic group	4.0	0.1	27.0	0.5	10.0	0.2	9.0	0.4	50.0	0.3
	Not Stated	2007.0	56.8	392.0	6.7	1041.0	21.4	119.0	5.2	3559.0	21.5
	Total	3534.0	21.3	5885.0	35.5	4863.0	29.3	2308.0	13.9	16590.0	100.0
Disability	Yes	123.0	3.5	620.0	10.5	318.0	6.5	205.0	8.9	1266.0	7.6
	No	3411.0	96.5	5265.0	89.5	4545.0	93.5	2103.0	91.1	15324.0	92.4
	Total	3534.0	21.3	5885.0	35.5	4863.0	29.3	2308.0	13.9	16590.0	100.0
Referred Problem	Mental & Behavioural Disorder due to alcohol use	0.0	0.0	3.0	0.1	3.0	0.1	3.0	0.1	9.0	0.1
	Bipolar Affective Disorder	2.0	0.1	11.0	0.2	1.0	0.0	2.0	0.1	16.0	0.1
	Depressive Episode	697.0	19.7	1517.0	25.8	1198.0	24.6	437.0	18.9	3849.0	23.2
	Recurrent Depressive Episode	328.0	9.3	318.0	5.4	295.0	6.1	187.0	8.1	1128.0	6.8
	Dysthymia		0.0	8.0	0.1	1.0	0.0	3.0	0.1	12.0	0.1
	Agoraphobia	25.0	0.7	27.0	0.5	37.0	0.8	14.0	0.6	103.0	0.6
	Social Phobias	24.0	0.7	55.0	0.9	52.0	1.1	45.0	1.9	176.0	1.1
	Specific Phobias	29.0	0.8	58.0	1.0	59.0	1.2	26.0	1.1	172.0	1.0
	Panic Disorder	164.0	4.6	259.0	4.4	186.0	3.8	142.0	6.2	751.0	4.5
	Generalized Anxiety Disorder	984.0	27.8	819.0	13.9	1107.0	22.8	382.0	16.6	3292.0	19.8
	Mixed Anxiety & Depressive Disorder	1061.0	30.0	2327.0	39.5	1408.0	29.0	634.0	27.5	5430.0	32.7
	Obsessive Compulsive Disorder	73.0	2.1	147.0	2.5	158.0	3.2	81.0	3.5	459.0	2.8

	Post Traumatic Stress Disorder	29.0	0.9	132.0	2.2	136.0	2.8	84.0	3.6	381.0	2.3
	Adjustment Disorder	1.0	0.0	36.0	0.6	18.0	0.4	24.0	1.0	79.0	0.5
	Somatoform Disorder	1.0	0.0	22.0	0.4	12.0	0.2	13.0	0.6	48.0	0.3
	Hypochondriac Disorder	3.0	0.1	5.0	0.1	4.0	0.1		0.0	12.0	0.1
	Eating Disorder	17.0	5.0	24.0	0.4	31.0	0.6	16.0	0.7	88.0	0.5
	Other Mental Disorder	84.0	2.4	101.0	1.7	106.0	2.2	213.0	9.2	504.0	3.0
	Disappearance or Death of Family Member	1.0	0.0	16.0	0.3	4.0	0.1	2.0	0.1	23.0	0.1
	Not Stated	11.0	0.3	0.0	0.0	47.0	1.0		0.0	58.0	0.3
	Total	3534.0	21.3	5885.0	35.5	4863.0	29.3	2308.0	13.9	16590.0	100.0
Employment Status First	Employed	1626.0	46.0	2673.0	45.4	2397.0	49.3	1191.0	51.6	7887.0	47.5
	Unemployed & Seeking Work	864.0	24.4	1071.0	18.2	1258.0	25.9	481.0	20.8	3674.0	22.1
	Students Not Seeking Work	260.0	7.4	526.0	8.9	195.0	4.0	100.0	4.3	1081.0	6.5
	Sick or Disabled	353.0	10.0	929.0	15.8	350.0	7.2	217.0	9.4	1849.0	11.1
	Homemaker	169.0	4.8	273.0	4.6	308.0	6.3	148.0	6.4	898.0	5.4
	No Benefits, Not Working or Seeking Work	2.0	0.1	27.0	0.5	5.0	0.1	1.0	0.0	35.0	0.2
	Voluntary Work	1.0	0.0	6.0	0.1		0.0		0.0	7.0	0.0
	Retired	224.0	6.3	356.0	6.0	316.0	6.5	153.0	6.6	1049.0	6.3
	Not Stated	35.0	1.0	24.0	0.4	34.0	0.7	17.0	0.7	110.0	0.7
	Total	3534.0	21.3	5885.0	35.5	4863.0	29.3	2308.0	13.9	16590.0	100.0

Appendix 5: Initial Scores at first session by years

		Year One		Year Two		Year Three		Year Four		Total	
PHQ First Score	Minimal	142	7.5%	227	5.3%	287	5.3%	271	5.3%	927	5.5%
	Mild	310	16.4%	554	12.9%	697	12.9%	718	14.0%	2279	13.6%
	Moderate	456	24.1%	959	22.3%	1206	22.4%	1144	22.2%	3765	22.5%
	Moderate to severe	544	28.7%	1163	27.1%	1556	28.8%	1497	29.1%	4760	28.5%
	Severe	441	23.3%	1388	32.3%	1648	30.6%	1515	29.4%	4992	29.9%
	Total	1893	11.3%	4291	25.7%	5394	32.3%	5145	30.8%	16723	100.0%
GAD First Score	Minimal	127	6.7%	220	5.1%	210	3.9%	212	4.1%	769	4.6%
	Mild	366	19.3%	668	15.6%	796	14.8%	848	16.5%	2678	16.0%
	Moderate	547	28.9%	1162	27.1%	1534	28.4%	1454	28.3%	4697	28.1%
	Severe	852	45.0%	2240	52.2%	2851	52.9%	2625	51.0%	8568	51.2%
	Not Recorded	1	0.1%	1	0.0%	3	0.1%	6	0.1%	11	0.1%
	Total	1893	11.3%	4291	25.7%	5394	32.3%	5145	30.8%	16723	100.0%
W&SAS First Score	Mild Functional Impairment	438	23.1%	792	18.5%	978	18.1%	953	18.5%	3161	18.9%
	Moderately Severe Functional Impairment	614	32.4%	1471	34.3%	1817	33.7%	1669	32.4%	5571	33.3%
	Severe Functional Impairment	814	43.0%	2020	47.1%	2591	48.0%	2515	48.9%	7940	47.5%
	Not Stated	27	1.4%	8	0.2%	8	0.1%	8	0.2%	51	0.3%
	Total	1893	11.3%	4291	25.7%	5394	32.3%	5145	30.8%	16723	100.0%

Appendix 6: Descriptive statistics for initial scores and outcomes

	Year One		Year Two		Year Three		Year Four		Total	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
PHQ First	14.3	6.4	15.7	6.3	15.6	6.2	15.4	6.2	15.4	6.3
PHQ Last	9.7	7.4	10.1	7.4	10.0	7.3	9.1	7.1	9.7	7.3
GAD First	13.1	5.3	13.9	5.1	14.2	5.0	14.0	5.0	13.9	5.1
GAD Last	8.6	6.3	9.0	6.4	9.0	6.3	8.2	6.1	8.7	6.3
W&SAS First	18.6	9.9	19.7	9.4	20.1	9.5	20.2	9.6	19.8	9.6
W&SAS Last	12.6	10.2	13.6	10.4	14.1	10.5	13.2	10.2	13.5	10.4

	Year One		Year Two		Year Three		Year Four		Total	
	Median	IR	Median	IR	Median	IR	Median	IR	Median	IR
PHQ First	15	9	16	10	16	10	16	9	16	9
PHQ Last	8	12	9	12	9	12	7	11	8	11
GAD First	14	9	15	8	15	7	15	7	15	8
GAD Last	7	11	8	10	8	10	7	10	7	11
W&SAS First	20	16	21	14	20	14	19	15	20	14
W&SAS Last	12	17	12	16	12	16	10	16	12	16

Appendix 7: Discharge reason analysis

By year

	Not Suitable	Referred On	Declined	Completed	Dropped Out	Not Known	Total
Year One	72	117	158	990	556	0	1893
%	3.80%	6.20%	8.30%	52.30%	29.40%	0.00%	100%
Year Two	139	263	261	2555	930	143	4291
%	3.20%	6.10%	6.10%	59.50%	21.70%	3.30%	100.00%
Year Three	126	415	372	3222	1062	197	5394
%	2.30%	7.70%	6.90%	59.70%	19.70%	3.70%	100.00%
Year Four	117	251	306	3289	1005	177	5145
%	2.30%	4.90%	5.90%	63.90%	19.50%	3.40%	100.00%

By Cohort

		Not Suitable	Referred On	Declined	Completed	Dropped Out	Not Known	Total
Model	Allocated	211	380	419	3545	1486	143	6184
	%	3.40%	6.10%	6.80%	57.30%	24.00%	2.30%	100%
	Progressive	243	666	678	6511	2067	374	10539
	%	2.30%	6.30%	6.40%	61.80%	19.60%	3.50%	100.00%
North	Allocated	88	203	229	2030	906	78	3534
	%	2.50%	5.70%	6.50%	57.40%	25.60%	2.20%	100.00%
	Progressive	155	363	354	3656	1169	188	5885
	%	2.60%	6.20%	6.00%	62.10%	19.90%	3.20%	100.00%
South	Allocated	157	352	353	2851	995	155	4863
	%	3.20%	7.20%	7.30%	58.60%	20.50%	3.20%	100.00%
	Progressive	48	154	154	1437	456	92	2341
	%	2.10%	6.60%	6.60%	61.40%	19.50%	3.90%	100.00%

By Model & Step

		Not Suitable	Referred On	Declined	Completed	Dropped Out	Total
Allocated	Step 2 Only	80	157	222	1896	758	3113
	%	2.57%	5.04%	7.13%	60.91%	24.35%	100.00%
	Step 3 Only	74	114	78	879	302	1447
	%	5.11%	7.88%	5.39%	60.75%	20.87%	100.00%
	Stepped Up	32	67	47	378	129	653
	%	4.90%	10.26%	7.20%	57.89%	19.75%	100.00%
	Stepped Down	2	4	7	64	16	93
	%	2.15%	4.30%	7.53%	68.82%	17.20%	100.00%
Progressive	Step 2 Only	126	302	484	3844	1482	6238
	%	2.02%	4.84%	7.76%	61.62%	23.76%	100.00%
	Step 3 Only	29	99	45	725	148	1046
	%	2.77%	9.46%	4.30%	69.31%	14.15%	100.00%
	Stepped Up	68	227	96	1568	309	2268
	%	3.00%	10.01%	4.23%	69.14%	13.62%	100.00%
	Stepped Down	1	12	9	106	29	157
	%	0.64%	7.64%	5.73%	67.52%	18.47%	100.00%

Appendix 8: Cross tabulation results comparing age against other variables

	Age	16-17	18-24	25-34	35-44	45-54	55-64	65-74	75-84	85-94	95+
Gender	Male	0.0%	30.9%	32.8%	36.5%	41.9%	41.9%	39.4%	30.9%	48.6%	0.0%
	Female	100.0%	69.1%	67.2%	63.5%	58.1%	58.1%	60.6%	69.1%	51.4%	100.0%
Disability	Yes	0.0%	3.0%	3.6%	5.4%	9.5%	13.5%	15.9%	19.3%	40.0%	100.0%
	No	100.0%	97.0%	96.4%	94.6%	90.5%	86.5%	84.1%	80.7%	60.0%	0.0%
Employment	Employed	0.0%	26.3%	53.7%	58.7%	54.2%	41.0%	15.9%	2.2%	0.0%	0.0%
	Unemployed	33.3%	24.1%	23.3%	21.6%	24.3%	25.5%	4.8%	0.0%	0.0%	0.0%
	Students	66.7%	43.0%	7.7%	1.6%	0.6%	0.0%	0.2%	0.0%	0.0%	0.0%
	Sick/Disabled	0.0%	3.3%	7.1%	10.7%	15.9%	19.7%	5.2%	0.0%	0.0%	0.0%
	Homemaker	0.0%	3.1%	8.0%	7.0%	4.1%	4.4%	1.3%	0.0%	0.0%	0.0%
	No Benefits	0.0%	0.2%	0.1%	0.2%	0.4%	0.3%	0.2%	0.0%	0.0%	0.0%
	Voluntary Work	0.0%	0.1%	0.0%	0.1%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%
	Retired	0.0%	0.0%	0.1%	0.1%	0.7%	9.1%	72.3%	97.8%	100.0%	100.0%

Appendix 9: Cross tabulation results comparing gender against other variables

	Gender	Male	Female
Age	16-17	0.0%	0.0%
	18-24	8.3%	11.0%
	25-34	20.7%	25.1%
	35-44	22.7%	23.4%
	45-54	24.9%	20.4%
	55-64	17.0%	14.0%
	65-74	5.2%	4.8%
	75-84	0.9%	1.2%
	85-94	0.3%	0.2%
	95+	0.0%	0.0%
Disability	Yes	9.1%	6.7%
	No	90.9%	93.3%
Employment	Employed	46.0%	48.8%
	Unemployed	28.0%	19.0%
	Students	5.2%	7.5%
	Sick/Disabled	13.3%	10.0%
	Homemaker	1.3%	7.8%
	No Benefits	0.3%	0.2%
	Voluntary Work	0.1%	0.0%
	Retired	5.9%	6.7%

Appendix 10: Cross tabulation results comparing disability against other variables

	Disability	Yes	No
Age	16-17	0.0%	0.0%
	18-24	3.9%	10.4%
	25-34	11.2%	24.5%
	35-44	16.5%	23.7%
	45-54	27.6%	21.6%
	55-64	26.6%	14.1%
	65-74	10.3%	4.5%
	75-84	2.7%	0.9%
	85-94	1.1%	0.1%
	95+	0.1%	0.0%
Gender	Male	44.5%	36.6%
	Female	55.5%	63.4%
Employment	Employed	17.0%	50.3%
	Unemployed	29.1%	21.8%
	Students	2.2%	7.0%
	Sick/Disabled	31.6%	9.5%
	Homemaker	3.2%	5.6%
	No Benefits	0.6%	0.2%
	Voluntary Work	0.1%	0.0%
	Retired	16.2%	5.6%

Appendix 11: Cross tabulation results – Age v Gender

Crosstab					
			gender		Total
			Male	Female	
Age	16-17	Count	0	3	3
		% within age	0.00%	100.00%	100.00%
		% within gender	0.00%	0.00%	0.00%
	18-24	Count	514	1150	1664
		% within age	30.90%	69.10%	100.00%
		% within gender	8.30%	11.00%	10.00%
	25-34	Count	1289	2641	3930
		% within age	32.80%	67.20%	100.00%
		% within gender	20.70%	25.10%	23.50%
	35-44	Count	1411	2454	3865
		% within age	36.50%	63.50%	100.00%
		% within gender	22.70%	23.40%	23.10%
	45-54	Count	1549	2145	3694
		% within age	41.90%	58.10%	100.00%
		% within gender	24.90%	20.40%	22.10%
	55-64	Count	1056	1465	2521
		% within age	41.90%	58.10%	100.00%
		% within gender	17.00%	14.00%	15.10%
	65-74	Count	325	499	824
		% within age	39.40%	60.60%	100.00%
		% within gender	5.20%	4.80%	4.90%
	75-84	Count	56	125	181
		% within age	30.90%	69.10%	100.00%
		% within gender	0.90%	1.20%	1.10%
	85-94	Count	17	18	35
		% within age	48.60%	51.40%	100.00%
		% within gender	0.30%	0.20%	0.20%
	95+	Count	0	1	1
		% within age	0.00%	100.00%	100.00%
		% within gender	0.00%	0.00%	0.00%
Total		Count	6217	10501	16718
		% within age	37.20%	62.80%	100.00%
		% within gender	100.00%	100.00%	100.00%

Appendix 12: Chi Square results - Age v Gender

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	130.013[a]	9	0
Likelihood Ratio	131.803	9	0
Linear-by-Linear Association	85.357	1	0
N of Valid Cases	16718		

a. 4 cells (20.0%) have expected count less than 5. The minimum expected count is .37.

Appendix 13: Cross tabulation results – Age v Employment

Crosstab										
			Employment Status First							Total
			Employed	Unemployed and Seeking Work	FT students or part students not seeking work	Sick, disabled, incapacity ben, income support, employ & support allowance	Homemaker	No benefits not working not seeking	Voluntary work, not working, not seeking	Retired	
Age	16-17	Count	0	1	2	0	0	0	0	0	3
		%within age	0.00%	33.30%	66.70%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%
		%within Employment Status First	0.00%	0.00%	0.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	18-24	Count	435	400	712	55	51	3	1	0	1657
		%within age	26.30%	24.10%	43.00%	3.30%	3.10%	0.20%	0.10%	0.00%	100.00%
		%within Employment Status First	5.50%	10.80%	64.70%	3.00%	5.70%	8.10%	14.30%	0.00%	10.00%
	25-34	Count	2095	908	299	277	312	4	1	3	3899
		%within age	53.70%	23.30%	7.70%	7.10%	8.00%	0.10%	0.00%	0.10%	100.00%
		%within Employment Status First	26.40%	24.50%	27.20%	14.90%	34.60%	10.80%	14.30%	0.30%	23.50%
	35-44	Count	2259	831	61	413	270	7	3	2	3846
		%within age	58.70%	21.60%	1.60%	10.70%	7.00%	0.20%	0.10%	0.10%	100.00%
		%within Employment Status First	28.50%	22.40%	5.50%	22.20%	29.90%	18.90%	42.90%	0.20%	23.20%
	45-54	Count	1987	890	23	582	149	13	0	24	3668
		%within age	54.20%	24.30%	0.60%	15.90%	4.10%	0.40%	0.00%	0.70%	100.00%
		%within Employment Status First	25.00%	24.00%	2.10%	31.20%	16.50%	35.10%	0.00%	2.30%	22.10%
	55-64	Count	1026	638	1	493	109	8	2	227	2504
		%within age	41.00%	25.50%	0.00%	19.70%	4.40%	0.30%	0.10%	9.10%	100.00%
		%within Employment Status First	12.90%	17.20%	0.10%	26.50%	12.10%	21.60%	28.60%	21.40%	15.10%
	65-74	Count	130	39	2	43	11	2	0	593	820
		%within age	15.90%	4.80%	0.20%	5.20%	1.30%	0.20%	0.00%	72.30%	100.00%
		%within Employment Status First	1.60%	1.10%	0.20%	2.30%	1.20%	5.40%	0.00%	55.90%	4.90%
	75-84	Count	4	0	0	0	0	0	0	176	180
		%within age	2.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	97.80%	100.00%
		%within Employment Status First	0.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	16.60%	1.10%
	85-94	Count	0	0	0	0	0	0	0	35	35
		%within age	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	100.00%
		%within Employment Status First	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.30%	0.20%
	95+	Count	0	0	0	0	0	0	0	1	1
		%within age	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	100.00%
		%within Employment Status First	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.10%	0.00%
Total		Count	7936	3707	1100	1863	902	37	7	1061	16613
		%within age	47.80%	22.30%	6.60%	11.20%	5.40%	0.20%	0.00%	6.40%	100.00%
		%within Employment Status First	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

Appendix 14: Chi Square results - Age v Employment

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	14520.994a	63	0
Likelihood Ratio	8201.987	63	0
Linear-by-Linear Association	2184.507	1	0
N of Valid Cases	16613		

a. 32 cells (40.0%) have expected count less than 5. The minimum expected count is .00.

Appendix 15: Cross tabulation results – Age v Disability

Crosstab					
			Disability		Total
			Yes	No	
Age	16-17	Count	0	3	3
		% within age	0.00%	100.00%	100.00%
		% within disability	0.00%	0.00%	0.00%
	18-24	Count	50	1614	1664
		% within age	3.00%	97.00%	100.00%
		% within disability	3.90%	10.40%	10.00%
	25-34	Count	143	3790	3933
		% within age	3.60%	96.40%	100.00%
		% within disability	11.20%	24.50%	23.50%
	35-44	Count	210	3655	3865
		% within age	5.40%	94.60%	100.00%
		% within disability	16.50%	23.70%	23.10%
	45-54	Count	352	3342	3694
		% within age	9.50%	90.50%	100.00%
		% within disability	27.60%	21.60%	22.10%
	55-64	Count	340	2182	2522
		% within age	13.50%	86.50%	100.00%
		% within disability	26.60%	14.10%	15.10%
	65-74	Count	131	694	825
		% within age	15.90%	84.10%	100.00%
		% within disability	10.30%	4.50%	4.90%
	75-84	Count	35	146	181
		% within age	19.30%	80.70%	100.00%
		% within disability	2.70%	0.90%	1.10%
	85-94	Count	14	21	35
		% within age	40.00%	60.00%	100.00%
		% within disability	1.10%	0.10%	0.20%
	95+	Count	1	0	1
		% within age	100.00%	0.00%	100.00%
		% within disability	0.10%	0.00%	0.00%
Total		Count	1276	15447	16723
		% within age	7.60%	92.40%	100.00%
		% within disability	100.00%	100.00%	100.00%

Appendix 16: Chi Square Results - Age v Disability

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	486.636a	9	0
Likelihood Ratio	443.304	9	0
Linear-by-Linear Association	438.157	1	0
N of Valid Cases	16723		

a. 5 cells (25.0%) have expected count less than 5. The minimum expected count is .08.

Appendix 17: Cross tabulation results – Gender v Disability

Crosstab					
			Disability		Total
			Yes	No	
Gender	Male	Count	568	5649	6217
		% within gender	9.10%	90.90%	100.00%
		% within disability	44.50%	36.60%	37.20%
	Female	Count	708	9793	10501
		% within gender	6.70%	93.30%	100.00%
		% within disability	55.50%	63.40%	62.80%
Total		Count	1276	15442	16718
		% within gender	7.60%	92.40%	100.00%
		% within disability	100.00%	100.00%	100.00%

Appendix 18: Chi Square results – Gender v Disability

Chi-Square Tests					
	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	31.747[a]	1	0		
Continuity Correction[b]	31.408	1	0		
Likelihood Ratio	31.1	1	0		
Fisher's Exact Test				0	0
Linear-by-Linear Association	31.745	1	0		
N of Valid Cases	16718				
a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 474.51.					
b. Computed only for a 2x2 table					

Appendix 19: Cross tabulation results – Gender v Employment

Crosstab					
			Gender		Total
			Male	Female	
Employment Status First	Employed	Count	2836	5098	7934
		% within Employment Status First	35.70%	64.30%	100.00%
		% within gender	46.00%	48.80%	47.80%
	Unemployed and Seeking Work	Count	1726	1980	3706
		% within Employment Status First	46.60%	53.40%	100.00%
		% within gender	28.00%	19.00%	22.30%
	ft students or part students not seeking work	Count	321	779	1100
		% within Employment Status First	29.20%	70.80%	100.00%
		% within gender	5.20%	7.50%	6.60%
	sick, disabled, incapacity ben, income support, employ & support allowance	Count	822	1041	1863
		% within Employment Status First	44.10%	55.90%	100.00%
		% within gender	13.30%	10.00%	11.20%
	homemaker	Count	83	817	900
		% within Employment Status First	9.20%	90.80%	100.00%
		% within gender	1.30%	7.80%	5.40%
	no benefits not working not seeking	Count	16	21	37
		% within Employment Status First	43.20%	56.80%	100.00%
		% within gender	0.30%	0.20%	0.20%
	vol work not working not seeking	Count	4	3	7
		% within Employment Status First	57.10%	42.90%	100.00%
		% within gender	0.10%	0.00%	0.00%
	retired	Count	362	699	1061
		% within Employment Status First	34.10%	65.90%	100.00%
		% within gender	5.90%	6.70%	6.40%
Total		Count	6170	10438	16608
		% within Employment Status First	37.20%	62.80%	100.00%
		% within gender	100.00%	100.00%	100.00%

Appendix 20: Chi Square results – Gender v Employment

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	522.950[a]	7	0
Likelihood Ratio	588.207	7	0
Linear-by-Linear Association	29.294	1	0
N of Valid Cases	16608		

a. 2 cells (12.5%) have expected count less than 5. The minimum expected count is 2.60.

Appendix 21: Cross tabulation results – Disability v Employment

Crosstab					
			disability		Total
			Yes	No	
Employment Status First	Employed	Count	215	7721	7936
		% within Employment Status First	2.70%	97.30%	100.00%
		% within disability	17.00%	50.30%	47.80%
	Unemployed and Seeking Work	Count	369	3338	3707
		% within Employment Status First	10.00%	90.00%	100.00%
		% within disability	29.10%	21.80%	22.30%
	ft students or part students not seeking work	Count	28	1072	1100
		% within Employment Status First	2.50%	97.50%	100.00%
		% within disability	2.20%	7.00%	6.60%
	sick, disabled, incapacity ben, income support, employ & support allowance	Count	401	1462	1863
		% within Employment Status First	21.50%	78.50%	100.00%
		% within disability	31.60%	9.50%	11.20%
	homemaker	Count	40	862	902
		% within Employment Status First	4.40%	95.60%	100.00%
		% within disability	3.20%	5.60%	5.40%
	no benefits not working not seeking	Count	8	29	37
		% within Employment Status First	21.60%	78.40%	100.00%
		% within disability	0.60%	0.20%	0.20%
	vol work not working not seeking	Count	1	6	7
		% within Employment Status First	14.30%	85.70%	100.00%
		% within disability	0.10%	0.00%	0.00%
	retired	Count	205	856	1061
		% within Employment Status First	19.30%	80.70%	100.00%
		% within disability	16.20%	5.60%	6.40%
Total		Count	1267	15346	16613
		% within Employment Status First	7.60%	92.40%	100.00%
		% within disability	100.00%	100.00%	100.00%

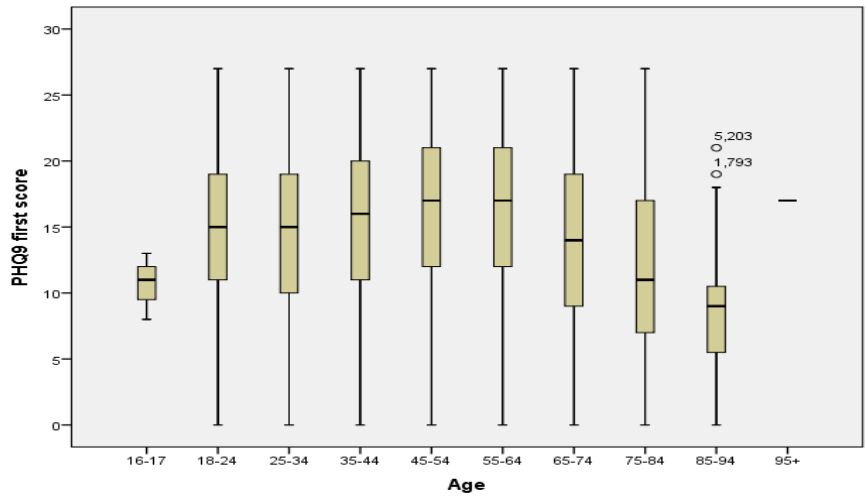
Appendix 22: Chi Square results – Disability v Employment

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1081.744[a]	7	0
Likelihood Ratio	962.826	7	0
Linear-by-Linear	544.437	1	0
N of Valid Cases	16613		

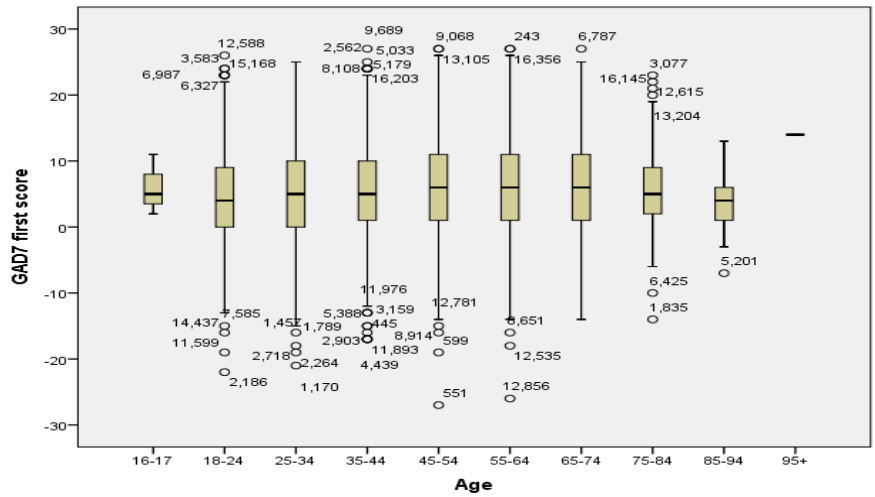
a. 2 cells (12.5%) have expected count less than 5. The minimum expected count is .53.

Appendix 23: Box plots comparing age against initial scores

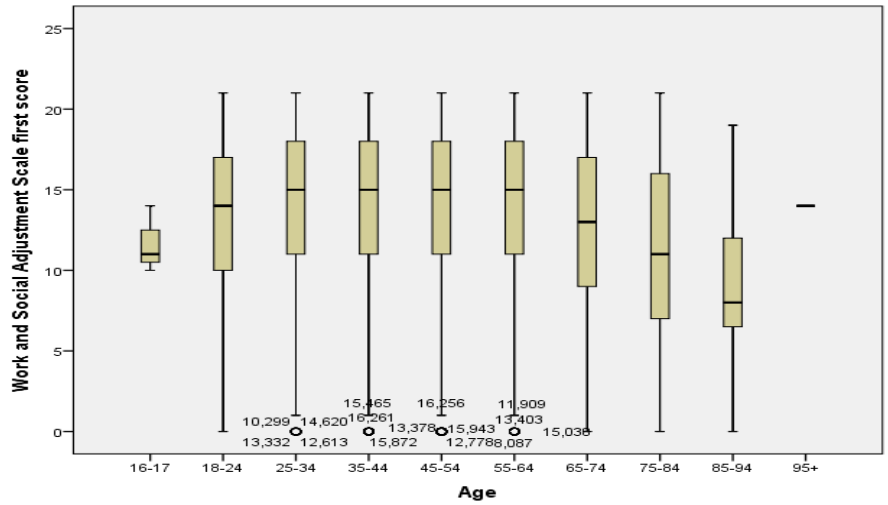
PHQ First Score



GAD First Score

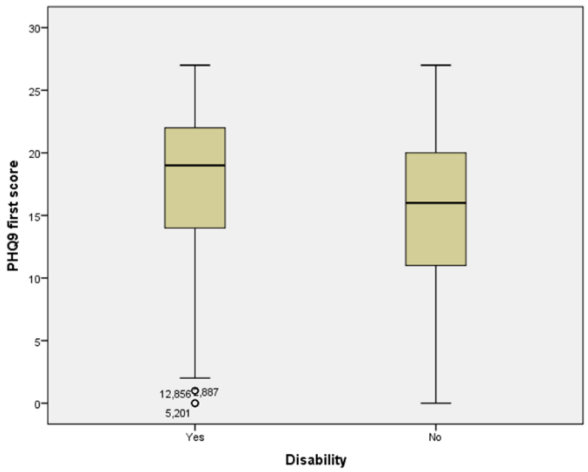


Work and Social Adjustment Scale First Score

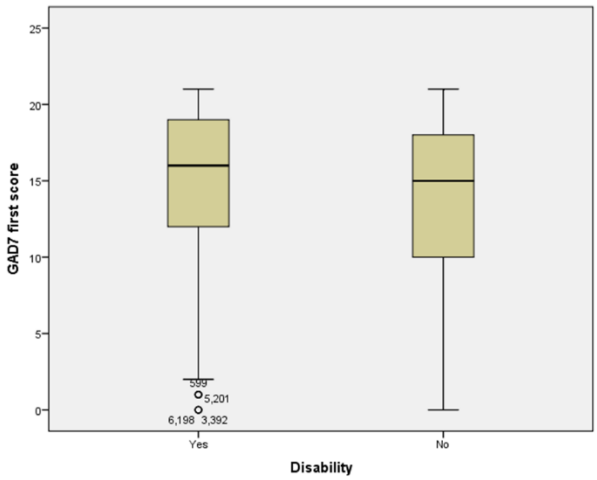


Appendix 24: Box plots comparing disability against initial scores

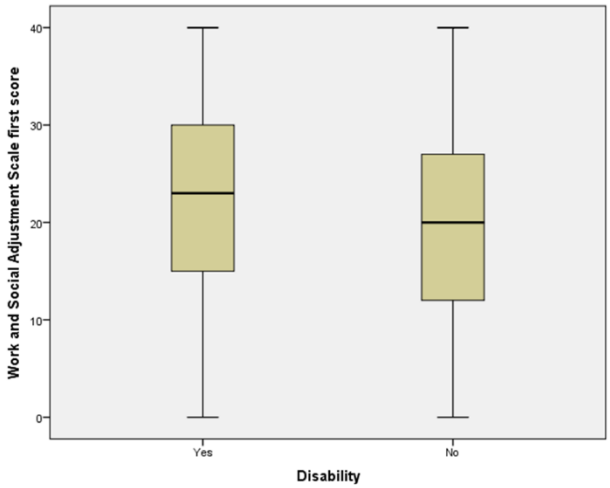
PHQ First Score



GAD First Score

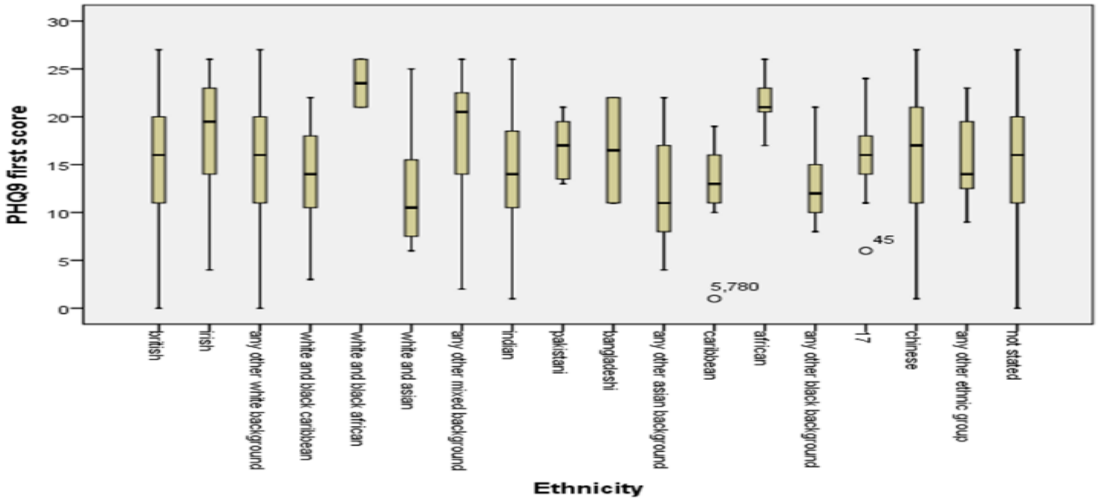


Work and Social Adjustment Scale First Score

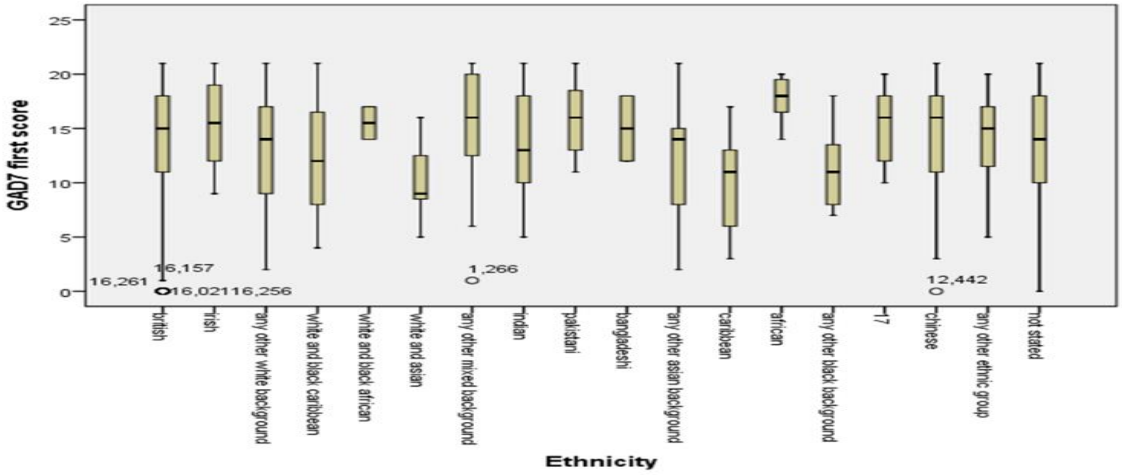


Appendix 25: Box plots comparing ethnicity against initial scores

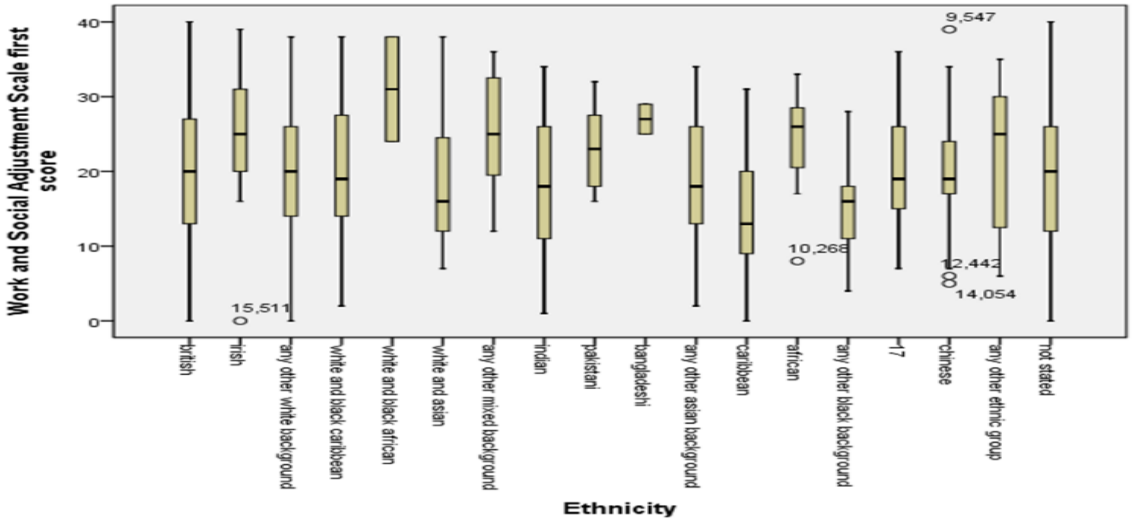
PHQ First Score



GAD First Score



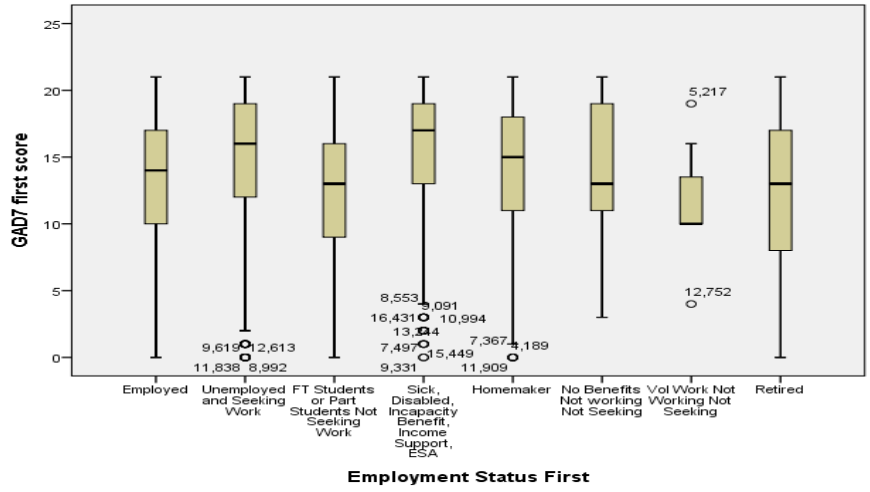
Work and Social Adjustment Scale First Score



Appendix 26: Box plots comparing first employment status against initial scores

PHQ First Score

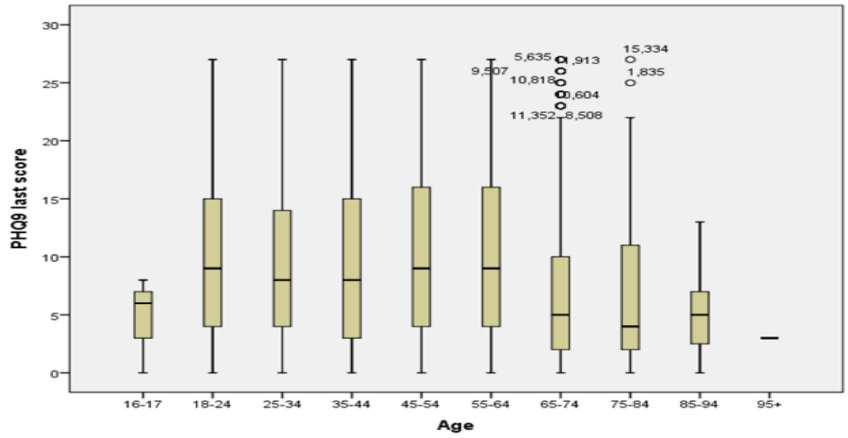
GAD First Score



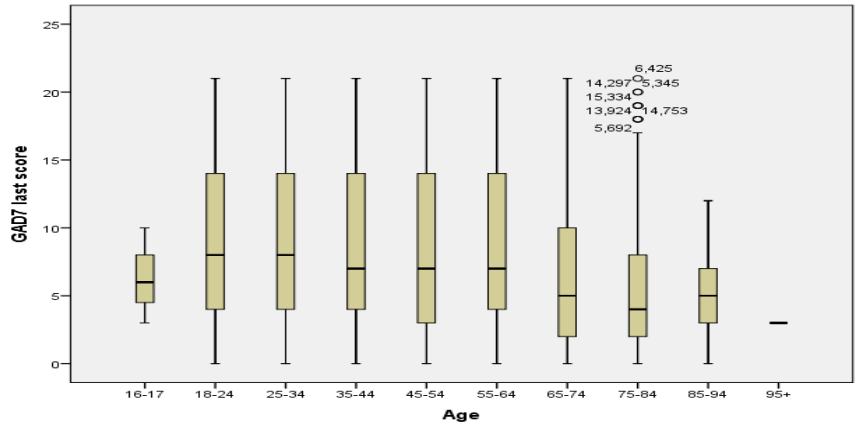
Work and Social Adjustment Scale First Score

Appendix 27: Box plots comparing age against last scores

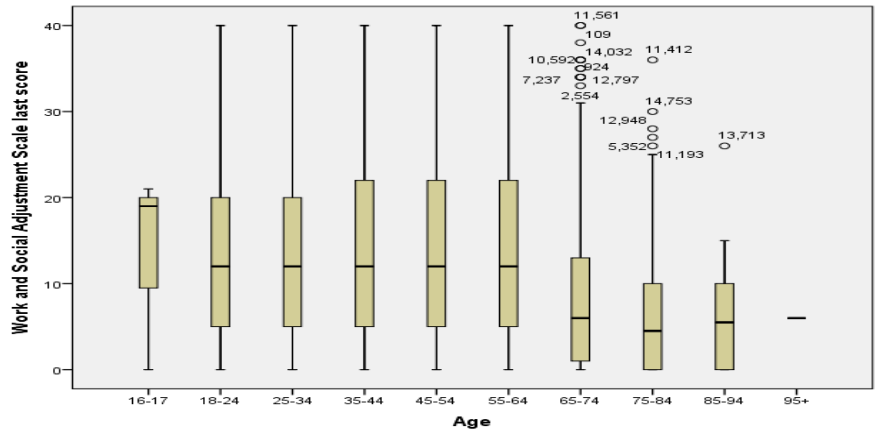
PHQ First Score



GAD First Score

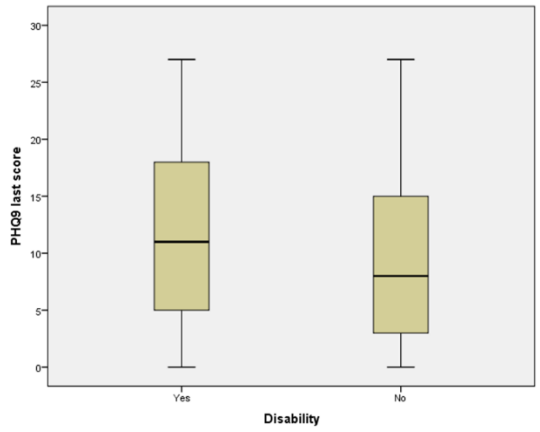


Work and Social Adjustment Scale First Score

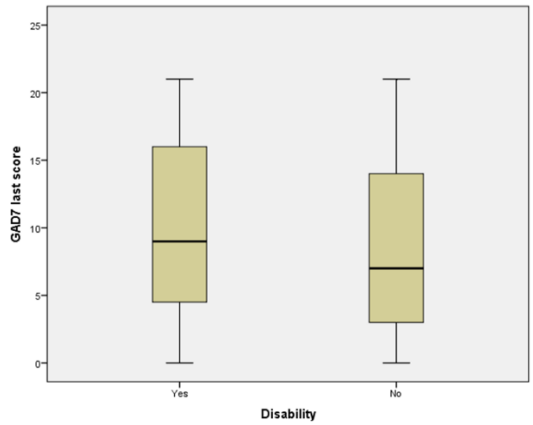


Appendix 28: Box plots comparing disability against last scores

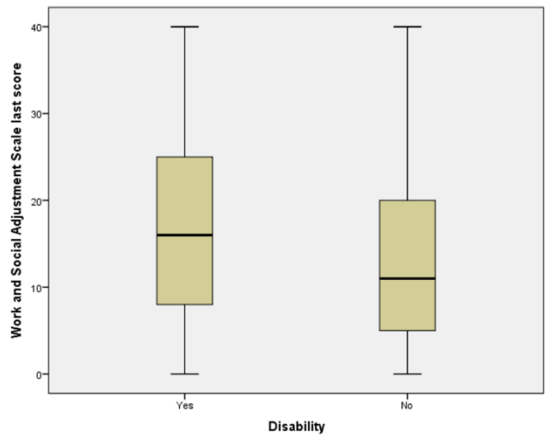
PHQ Last Score



GAD Last Score

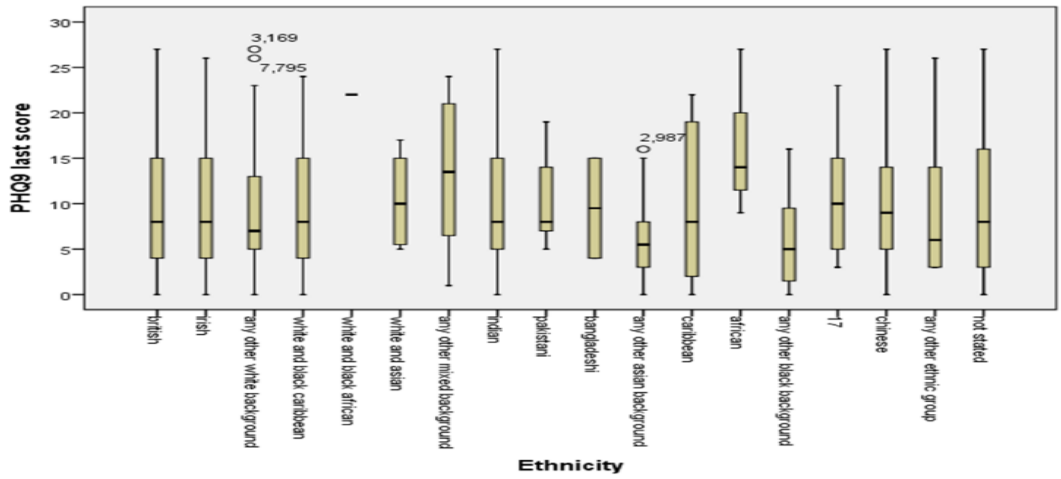


Work and Social Adjustment Scale Last Score

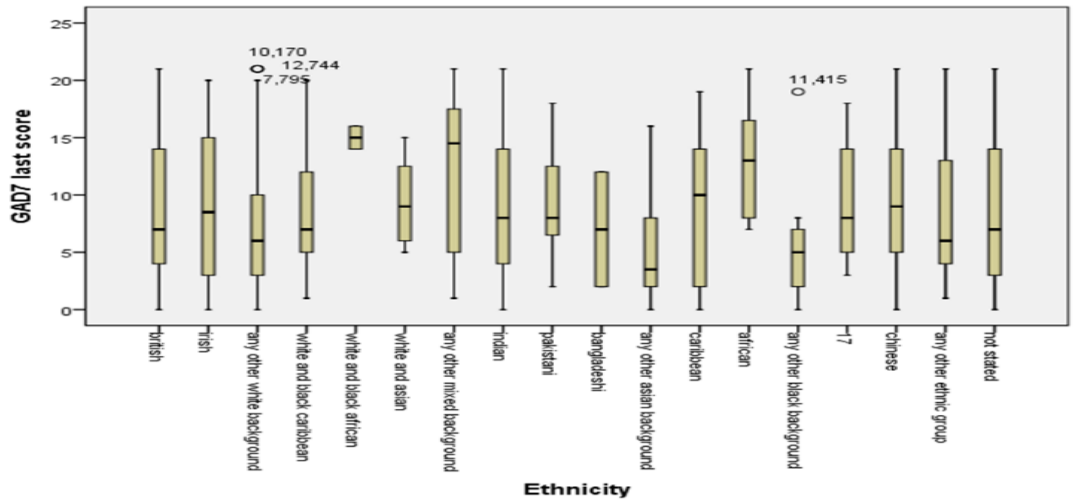


Appendix 29: Box plots comparing ethnicity against last scores

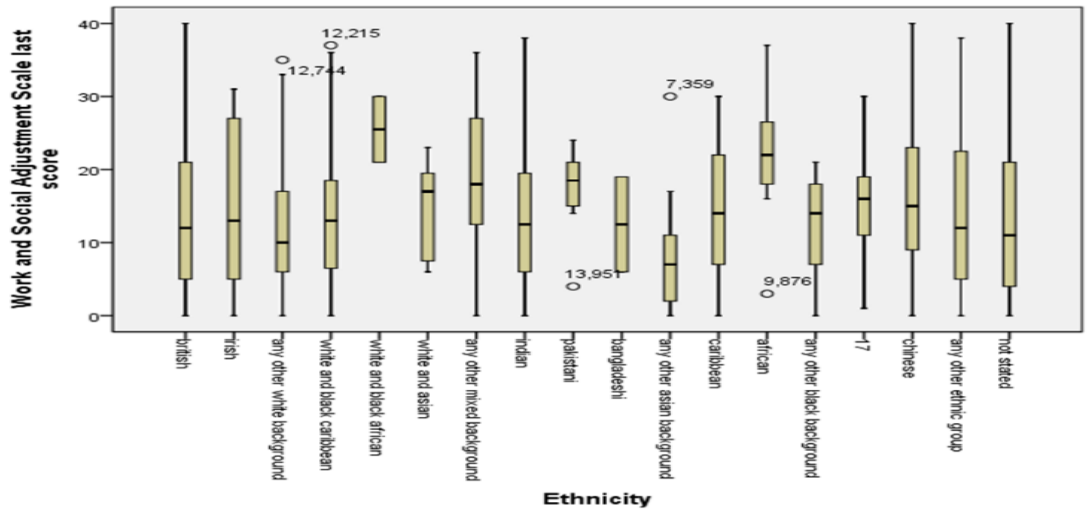
PHQ First Score



GAD First Score

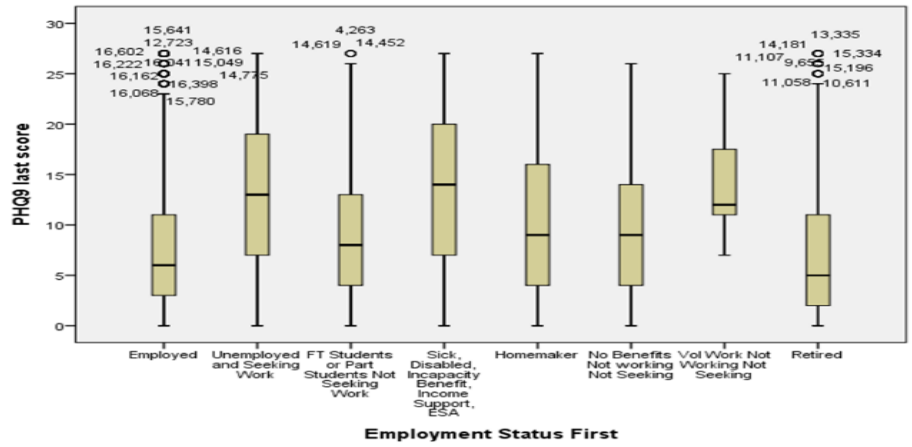


Work and Social Adjustment Scale First Score

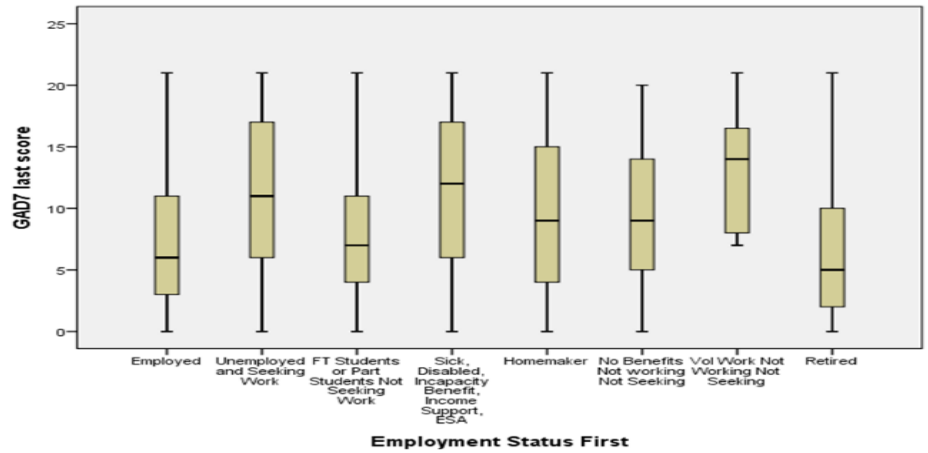


Appendix 30: Box plots comparing first employment status against last scores

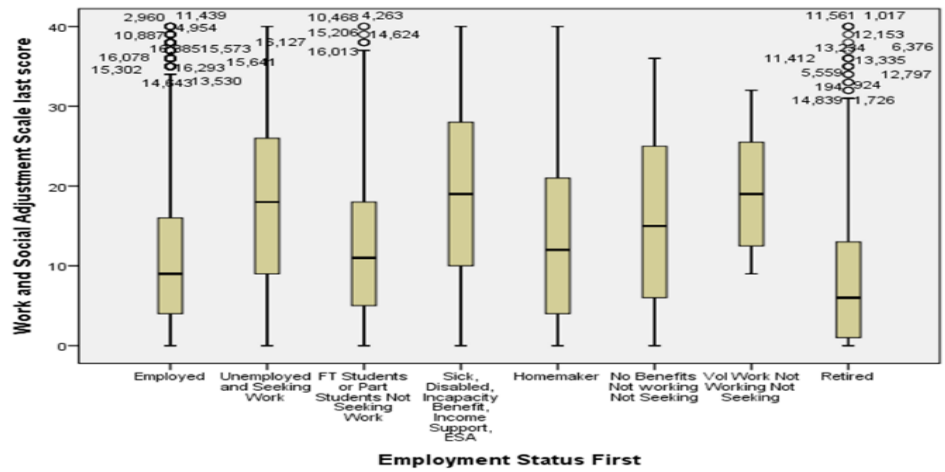
PHQ First Score



GAD First Score



Work and Social Adjustment Scale First Score



Appendix 31: Cross tabulation of model and outcome for whole service

Recovery * Model Crosstabulation

			Model		Total
			Allocated	Progressive	
Recovery	Recovered	Count	2323	4486	6809
		% within Model	41.40%	46.10%	44.40%
	Reliable Improvement	Count	1401	2360	3761
		% within Model	25.00%	24.30%	24.50%
	Non Recovered	Count	1570	2458	4028
		% within Model	28.00%	25.30%	26.30%
	Reliable Deterioration	Count	311	426	737
		% within Model	5.50%	4.40%	4.80%
Total		Count	5605	9730	15335
		% within Model	100.00%	100.00%	100.00%

Appendix 32: Chi square of model and outcome for whole service

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	38.552[a]	3	0
Likelihood Ratio	38.44	3	0
Linear-by-Linear Association	37.906	1	0
N of Valid Cases	15335		

Appendix 33: Cross tabulation of model on recovery v non recovery for whole service

Recovery Non Recovery * Model Crosstabulation

			Model		Total
			Allocated	Progressive	
Recovery Non Recovery	Recovered	Count	2323	4486	6809
		% within Model	41.40%	46.10%	44.40%
	Non Recovered	Count	3282	5244	8526
		% within Model	58.60%	53.90%	55.60%
Total		Count	5605	9730	15335
		% within Model	100.00%	100.00%	100.00%

Appendix 34: Chi square of model on recovery v non recovery for whole service

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	31.279[a]	1	0		
Continuity Correction[b]	31.091	1	0		
Likelihood Ratio	31.353	1	0		
Fisher's Exact Test				0	0
Linear-by-Linear Association	31.277	1	0		
N of Valid Cases	15335				

Appendix 35: Cross tabulation of model on reliable improvement v non recovery for whole service

Reliable Improvement Non Recovery * Model Crosstabulation

			Model		Total
			Allocated	Progressive	
Reliable Improvement Non Recovery	Reliable Improvement	Count	1401	2360	3761
		% within Model	42.70%	45.00%	44.10%
	Non Recovery	Count	1881	2884	4765
		% within Model	57.30%	55.00%	55.90%
Total		Count	3282	5244	8526
		% within Model	100.00%	100.00%	100.00%

Appendix 36: Chi square of model on reliable improvement v non recovery for whole service

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	4.394[a]	1	0.036		
Continuity Correction[b]	4.3	1	0.038		
Likelihood Ratio	4.398	1	0.036		
Fisher's Exact Test				0.037	0.019
Linear-by-Linear Association	4.393	1	0.036		
N of Valid Cases	8526				

Appendix 37: Cross tabulation of model on reliable deterioration v non recovery for whole service

RD V Non Recovery * Model Crosstabulation

			Model		Total
			Allocated	Progressive	
RD V Non Recovery	Reliable Deterioration	Count	311	426	737
		% within Model	9.50%	8.10%	8.60%
	Non Recovery	Count	2971	4818	7789
		% within Model	90.50%	91.90%	91.40%
Total		Count	3282	5244	8526
		% within Model	100.00%	100.00%	100.00%

Appendix 38: Chi square of model on reliable deterioration v non recovery for whole service

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	4.675[a]	1	0.031		
Continuity Correction[b]	4.505	1	0.034		
Likelihood Ratio	4.629	1	0.031		
Fisher's Exact Test				0.032	0.017
Linear-by-Linear Association	4.674	1	0.031		
N of Valid Cases	8526				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 283.70.

b. Computed only for a 2x2 table

Appendix 39: Cross tabulation on PHQ Moderate to Severe/Severe on model v outcomes for whole service

Crosstab

PHQ Severity				Model		Total
				Allocated	Progressive	
Severe	Recovery	Recovered	Count	456	924	1380
			% within Model	24.90%	29.30%	27.70%
		Reliable Improvement	Count	645	1143	1788
			% within Model	35.30%	36.20%	35.90%
		Non Recovered	Count	676	1022	1698
			% within Model	37.00%	32.40%	34.10%
		Reliable Deterioration	Count	51	69	120
			% within Model	2.80%	2.20%	2.40%
	Total		Count	1828	3158	4986
			% within Model	100.00%	100.00%	100.00%
Moderate to Severe	Recovery	Recovered	Count	653	1370	2023
			% within Model	38.30%	44.90%	42.60%
		Reliable Improvement	Count	495	778	1273
			% within Model	29.00%	25.50%	26.80%
		Non Recovered	Count	456	754	1210
			% within Model	26.70%	24.70%	25.50%
		Reliable Deterioration	Count	101	147	248
			% within Model	5.90%	4.80%	5.20%
	Total		Count	1705	3049	4754
			% within Model	100.00%	100.00%	100.00%
Total	Recovery	Recovered	Count	1109	2294	3403
			% within Model	31.40%	37.00%	34.90%
		Reliable Improvement	Count	1140	1921	3061
			% within Model	32.30%	30.90%	31.40%
		Non Recovered	Count	1132	1776	2908
			% within Model	32.00%	28.60%	29.90%
		Reliable Deterioration	Count	152	216	368
			% within Model	4.30%	3.50%	3.80%
	Total		Count	3533	6207	9740
			% within Model	100.00%	100.00%	100.00%

Appendix 40: Chi square on PHQ Moderate to Severe/Severe on model v outcomes for whole service

Chi-Square Tests

PHQ Severity		Value	df	Asymp. Sig. (2-sided)
Severe	Pearson Chi-Square	17.063[b]	3	0.001
	Likelihood Ratio	17.066	3	0.001
	Linear-by-Linear Association	16.983	1	0
	N of Valid Cases	4986		
Moderate to Severe	Pearson Chi-Square	20.649[c]	3	0
	Likelihood Ratio	20.714	3	0
	Linear-by-Linear Association	14.566	1	0
	N of Valid Cases	4754		
Total	Pearson Chi-Square	34.118[a]	3	0
	Likelihood Ratio	34.274	3	0
	Linear-by-Linear Association	31.926	1	0
	N of Valid Cases	9740		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 133.49.
b. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 44.00.
c. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 88.94.

Appendix 41: Cross tabulation on GAD Moderate/Severe on model v outcomes for whole service

Crosstab

GAD Severity				Model		Total
				Allocated	Progressive	
Severe	Recovery	Recovered	Count	964	1971	2935
			% within Model	31.20%	36.00%	34.30%
		Reliable Improvement	Count	1064	1827	2891
			% within Model	34.40%	33.40%	33.70%
		Non Recovered	Count	970	1558	2528
			% within Model	31.40%	28.50%	29.50%
		Reliable Deterioration	Count	93	120	213
			% within Model	3.00%	2.20%	2.50%
	Total		Count	3091	5476	8567
			% within Model	100.00%	100.00%	100.00%
Moderate	Recovery	Recovered	Count	854	1666	2520
			% within Model	50.00%	55.80%	53.70%
		Reliable Improvement	Count	284	474	758
			% within Model	16.60%	15.90%	16.10%
		Non Recovered	Count	440	650	1090
			% within Model	25.80%	21.80%	23.20%
		Reliable Deterioration	Count	130	196	326
			% within Model	7.60%	6.60%	6.90%
	Total		Count	1708	2986	4694
			% within Model	100.00%	100.00%	100.00%
Total	Recovery	Recovered	Count	1818	3637	5455
			% within Model	37.90%	43.00%	41.10%
		Reliable Improvement	Count	1348	2301	3649
			% within Model	28.10%	27.20%	27.50%
		Non Recovered	Count	1410	2208	3618
			% within Model	29.40%	26.10%	27.30%
		Reliable Deterioration	Count	223	316	539
			% within Model	4.60%	3.70%	4.10%
	Total		Count	4799	8462	13261
			% within Model	100.00%	100.00%	100.00%

Appendix 42: Chi square on GAD Moderate/Severe on model v outcomes for whole service

Chi-Square Tests

GAD Severity		Value	df	Asymp. Sig. (2-sided)
Severe	Pearson Chi-Square	25.034[b]	3	0
	Likelihood Ratio	25.035	3	0
	Linear-by-Linear Association	23.534	1	0
	N of Valid Cases	8567		
Moderate	Pearson Chi-Square	16.351[c]	3	0.001
	Likelihood Ratio	16.293	3	0.001
	Linear-by-Linear Association	15.042	1	0
	N of Valid Cases	4694		
Total	Pearson Chi-Square	38.647[a]	3	0
	Likelihood Ratio	38.64	3	0
	Linear-by-Linear Association	37.846	1	0
	N of Valid Cases	13261		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 195.06.
b. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 76.85.
c. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 118.62.

Appendix 43: Cross tabulation for age against model and recovery outcome for whole service

Recovery				Model		Total
				Allocated	Progressive	
Recovered	Age	16-17	Count	0	2	2
			% within Model	0.00%	0.00%	0.00%
		18-24	Count	136	488	624
			% within Model	5.90%	10.90%	9.20%
		25-34	Count	517	1011	1528
			% within Model	22.30%	22.50%	22.40%
		35-44	Count	531	1079	1610
			% within Model	22.90%	24.10%	23.60%
		45-54	Count	567	966	1533
			% within Model	24.40%	21.50%	22.50%
		55-64	Count	366	618	984
			% within Model	15.80%	13.80%	14.50%
		65-74	Count	176	253	429
			% within Model	7.60%	5.60%	6.30%
		75-84	Count	24	56	80
			% within Model	1.00%	1.20%	1.20%
		85-94	Count	6	12	18
			% within Model	0.30%	0.30%	0.30%
		95+	Count	0	1	1
			% within Model	0.00%	0.00%	0.00%
	Total		Count	2323	4486	6809
			% within Model	100.00%	100.00%	100.00%
Reliable Improvement	Age	18-24	Count	82	269	351
			% within Model	5.90%	11.40%	9.30%
		25-34	Count	357	624	981
			% within Model	25.50%	26.40%	26.10%
		35-44	Count	337	498	835
			% within Model	24.10%	21.10%	22.20%
		45-54	Count	331	493	824
			% within Model	23.60%	20.90%	21.90%
		55-64	Count	235	380	615
			% within Model	16.80%	16.10%	16.40%
		65-74	Count	50	73	123
			% within Model	3.60%	3.10%	3.30%
		75-84	Count	8	20	28
			% within Model	0.60%	0.80%	0.70%
		85-94	Count	1	3	4
			% within Model	0.10%	0.10%	0.10%
	Total		Count	1401	2360	3761
			% within Model	100.00%	100.00%	100.00%

Recovery				Model		Total
				Allocated	Progressive	
Non Recovered	Age	16-17	Count	0	1	1
			% within Model	0.00%	0.00%	0.00%
		18-24	Count	123	351	474
			% within Model	7.80%	14.30%	11.80%
		25-34	Count	395	566	961
			% within Model	25.20%	23.00%	23.90%
		35-44	Count	367	549	916
			% within Model	23.40%	22.30%	22.70%
		45-54	Count	360	541	901
			% within Model	22.90%	22.00%	22.40%
		55-64	Count	253	364	617
			% within Model	16.10%	14.80%	15.30%
		65-74	Count	60	68	128
			% within Model	3.80%	2.80%	3.20%
		75-84	Count	11	15	26
			% within Model	0.70%	0.60%	0.60%
		85-94	Count	1	3	4
			% within Model	0.10%	0.10%	0.10%
	Total		Count	1570	2458	4028
			% within Model	100.00%	100.00%	100.00%
Reliable Deterioration	Age	18-24	Count	22	65	87
			% within Model	7.10%	15.30%	11.80%
		25-34	Count	76	90	166
			% within Model	24.40%	21.10%	22.50%
		35-44	Count	88	102	190
			% within Model	28.30%	23.90%	25.80%
		45-54	Count	65	88	153
			% within Model	20.90%	20.70%	20.80%
		55-64	Count	45	62	107
			% within Model	14.50%	14.60%	14.50%
		65-74	Count	13	19	32
			% within Model	4.20%	4.50%	4.30%
		75-84	Count	2	0	2
			% within Model	0.60%	0.00%	0.30%
	Total		Count	311	426	737
			% within Model	100.00%	100.00%	100.00%

Recovery				Model		Total
				Allocated	Progressive	
Total	Age	16-17	Count	0	3	3
			% within Model	0.00%	0.00%	0.00%
		18-24	Count	363	1173	1536
			% within Model	6.50%	12.10%	10.00%
		25-34	Count	1345	2291	3636
			% within Model	24.00%	23.50%	23.70%
		35-44	Count	1323	2228	3551
			% within Model	23.60%	22.90%	23.20%
		45-54	Count	1323	2088	3411
			% within Model	23.60%	21.50%	22.20%
		55-64	Count	899	1424	2323
			% within Model	16.00%	14.60%	15.10%
		65-74	Count	299	413	712
			% within Model	5.30%	4.20%	4.60%
		75-84	Count	45	91	136
			% within Model	0.80%	0.90%	0.90%
		85-94	Count	8	18	26
			% within Model	0.10%	0.20%	0.20%
		95+	Count	0	1	1
			% within Model	0.00%	0.00%	0.00%
	Total		Count	5605	9730	15335
			% within Model	100.00%	100.00%	100.00%

Appendix 44: Chi square for age against model and recovery outcome for whole service

Chi-Square Tests

Recovery		Value	df	Asymp. Sig. (2-sided)
Recovered	Pearson Chi-Square	64.164[b]	9	0
	Likelihood Ratio	68.05	9	0
	Linear-by-Linear Association	35.532	1	0
	N of Valid Cases	6809		
Reliable Improvement	Pearson Chi-Square	37.743[c]	7	0
	Likelihood Ratio	39.81	7	0
	Linear-by-Linear Association	12.139	1	0
	N of Valid Cases	3761		
Non Recovered	Pearson Chi-Square	41.980[d]	8	0
	Likelihood Ratio	44.197	8	0
	Linear-by-Linear Association	15.464	1	0
	N of Valid Cases	4028		
Reliable Deterioration	Pearson Chi-Square	15.174[e]	6	0.019
	Likelihood Ratio	16.5	6	0.011
	Linear-by-Linear Association	2.1	1	0.147
	N of Valid Cases	737		
Total	Pearson Chi-Square	136.049[a]	9	0
	Likelihood Ratio	144.774	9	0
	Linear-by-Linear Association	59.057	1	0
	N of Valid Cases	15335		

Appendix 45: Cross tabulation for gender against model and recovery outcome for whole service

Crosstab

Recovery				Model		Total
				Allocated	Progressive	
Recovered	gender	Male	Count	841	1676	2517
			% within Model	36.20%	37.40%	37.00%
		Female	Count	1482	2809	4291
			% within Model	63.80%	62.60%	63.00%
	Total		Count	2323	4485	6808
			% within Model	100.00%	100.00%	100.00%
Reliable Improvement	gender	Male	Count	529	853	1382
			% within Model	37.80%	36.20%	36.80%
		Female	Count	872	1506	2378
			% within Model	62.20%	63.80%	63.20%
	Total		Count	1401	2359	3760
			% within Model	100.00%	100.00%	100.00%
Non Recovered	gender	Male	Count	602	939	1541
			% within Model	38.40%	38.20%	38.30%
		Female	Count	966	1519	2485
			% within Model	61.60%	61.80%	61.70%
	Total		Count	1568	2458	4026
			% within Model	100.00%	100.00%	100.00%
Reliable Deterioration	gender	Male	Count	112	144	256
			% within Model	36.00%	33.80%	34.70%
		Female	Count	199	282	481
			% within Model	64.00%	66.20%	65.30%
	Total		Count	311	426	737
			% within Model	100.00%	100.00%	100.00%
Total	gender	Male	Count	2084	3612	5696
			% within Model	37.20%	37.10%	37.20%
		Female	Count	3519	6116	9635
			% within Model	62.80%	62.90%	62.80%
	Total		Count	5603	9728	15331
			% within Model	100.00%	100.00%	100.00%

Appendix 46: Chi square for gender against model and recovery outcome for whole service

Chi-Square Tests

Recovery		Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Recovered	Pearson Chi-Square	.893[c]	1	0.345		
	Continuity Correction[b]	0.843	1	0.358		
	Likelihood Ratio	0.894	1	0.344		
	Fisher's Exact Test				0.354	0.179
	Linear-by-Linear Association	0.892	1	0.345		
	N of Valid Cases	6808				
Reliable Improvement	Pearson Chi-Square	.967[d]	1	0.325		
	Continuity Correction[b]	0.9	1	0.343		
	Likelihood Ratio	0.966	1	0.326		
	Fisher's Exact Test				0.328	0.171
	Linear-by-Linear Association	0.967	1	0.325		
	N of Valid Cases	3760				
Non Recovered	Pearson Chi-Square	.015[e]	1	0.903		
	Continuity Correction[b]	0.008	1	0.93		
	Likelihood Ratio	0.015	1	0.903		
	Fisher's Exact Test				0.921	0.465
	Linear-by-Linear Association	0.015	1	0.903		
	N of Valid Cases	4026				
Reliable Deterioration	Pearson Chi-Square	.387[f]	1	0.534		
	Continuity Correction[b]	0.296	1	0.586		
	Likelihood Ratio	0.387	1	0.534		
	Fisher's Exact Test				0.584	0.293
	Linear-by-Linear Association	0.387	1	0.534		
	N of Valid Cases	737				
Total	Pearson Chi-Square	.006[a]	1	0.937		
	Continuity Correction[b]	0.004	1	0.95		
	Likelihood Ratio	0.006	1	0.937		
	Fisher's Exact Test				0.945	0.475
	Linear-by-Linear Association	0.006	1	0.937		
	N of Valid Cases	15331				

Appendix 47: Cross tabulation for disability against model and recovery outcome for whole service

Crosstab

Recovery				Model		Total
				Allocated	Progressive	
Recovered	Disability	Yes	Count	81	365	446
			% within Model	3.50%	8.10%	6.60%
		No	Count	2242	4121	6363
			% within Model	96.50%	91.90%	93.40%
	Total		Count	2323	4486	6809
			% within Model	100.00%	100.00%	100.00%
Reliable Improvement	Disability	Yes	Count	56	276	332
			% within Model	4.00%	11.70%	8.80%
		No	Count	1345	2084	3429
			% within Model	96.00%	88.30%	91.20%
	Total		Count	1401	2360	3761
			% within Model	100.00%	100.00%	100.00%
Non Recovered	Disability	Yes	Count	75	302	377
			% within Model	4.80%	12.30%	9.40%
		No	Count	1495	2156	3651
			% within Model	95.20%	87.70%	90.60%
	Total		Count	1570	2458	4028
			% within Model	100.00%	100.00%	100.00%
Reliable Deterioration	Disability	Yes	Count	15	50	65
			% within Model	4.80%	11.70%	8.80%
		No	Count	296	376	672
			% within Model	95.20%	88.30%	91.20%
	Total		Count	311	426	737
			% within Model	100.00%	100.00%	100.00%
Total	Disability	Yes	Count	227	993	1220
			% within Model	4.00%	10.20%	8.00%
		No	Count	5378	8737	14115
			% within Model	96.00%	89.80%	92.00%
	Total		Count	5605	9730	15335
			% within Model	100.00%	100.00%	100.00%

Appendix 48: Chi square for disability against model and recovery outcome for whole service

Chi-Square Tests

Recovery		Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Recovered	Pearson Chi-Square	54.053[c]	1	0		
	Continuity Correction[b]	53.296	1	0		
	Likelihood Ratio	59.693	1	0		
	Fisher's Exact Test				0	0
	Linear-by-Linear Association	54.045	1	0		
	N of Valid Cases	6809				
Reliable Improvement	Pearson Chi-Square	64.726[d]	1	0		
	Continuity Correction[b]	63.773	1	0		
	Likelihood Ratio	72.21	1	0		
	Fisher's Exact Test				0	0
	Linear-by-Linear Association	64.709	1	0		
	N of Valid Cases	3761				
Non Recovered	Pearson Chi-Square	63.683[e]	1	0		
	Continuity Correction[b]	62.801	1	0		
	Likelihood Ratio	69.392	1	0		
	Fisher's Exact Test				0	0
	Linear-by-Linear Association	63.667	1	0		
	N of Valid Cases	4028				
Reliable Deterioration	Pearson Chi-Square	10.686[f]	1	0.001		
	Continuity Correction[b]	9.843	1	0.002		
	Likelihood Ratio	11.411	1	0.001		
	Fisher's Exact Test				0.001	0.001
	Linear-by-Linear Association	10.671	1	0.001		
	N of Valid Cases	737				
Total	Pearson Chi-Square	184.023[a]	1	0		
	Continuity Correction[b]	183.183	1	0		
	Likelihood Ratio	202.635	1	0		
	Fisher's Exact Test				0	0
	Linear-by-Linear Association	184.011	1	0		
	N of Valid Cases	15335				

Appendix 49: Cross tabulation for first employment status against model and recovery outcome for whole service

Recovery				Model		Total
				Allocated	Progressive	
Recovered	Employment Status First	Employed	Count	1348	2595	3943
			% within Model	58.60%	58.00%	58.20%
		Unemployed and Seeking Work	Count	419	596	1015
			% within Model	18.20%	13.30%	15.00%
		ft students or part students not seeking work	Count	145	318	463
			% within Model	6.30%	7.10%	6.80%
		sick, disabled, incapacity ben, income support, employ & support allowance	Count	120	386	506
			% within Model	5.20%	8.60%	7.50%
		homemaker	Count	107	217	324
			% within Model	4.70%	4.90%	4.80%
		no benefits not working not seeking	Count	0	14	14
			% within Model	0.00%	0.30%	0.20%
		vol work not working not seeking	Count	0	1	1
			% within Model	0.00%	0.00%	0.00%
		retired	Count	162	346	508
			% within Model	7.00%	7.70%	7.50%
	Total		Count	2301	4473	6774
			% within Model	100.00%	100.00%	100.00%
Reliable Improvement	Employment Status First	Employed	Count	603	950	1553
			% within Model	43.30%	40.40%	41.50%
		Unemployed and Seeking Work	Count	409	619	1028
			% within Model	29.40%	26.30%	27.50%
		ft students or part students not seeking work	Count	73	157	230
			% within Model	5.20%	6.70%	6.10%
		sick, disabled, incapacity ben, income support, employ & support allowance	Count	164	379	543
			% within Model	11.80%	16.10%	14.50%
		homemaker	Count	87	119	206
			% within Model	6.20%	5.10%	5.50%
		no benefits not working not seeking	Count	0	9	9
			% within Model	0.00%	0.40%	0.20%
		vol work not working not seeking	Count	0	1	1
			% within Model	0.00%	0.00%	0.00%
		retired	Count	57	116	173
			% within Model	4.10%	4.90%	4.60%
	Total		Count	1393	2350	3743
			% within Model	100.00%	100.00%	100.00%

Recovery				Model		Total
				Allocated	Progressive	
Non Recovered	Employment Status First	Employed	Count	541	835	1376
			% within Model	34.80%	34.10%	34.40%
		Unemployed and Seeking Work	Count	543	710	1253
			% within Model	34.90%	29.00%	31.30%
		ft students or part students not seeking work	Count	86	178	264
			% within Model	5.50%	7.30%	6.60%
		sick, disabled, incapacity ben, income support, employ & support allowance	Count	194	462	656
			% within Model	12.50%	18.90%	16.40%
		homemaker	Count	110	147	257
			% within Model	7.10%	6.00%	6.40%
		no benefits not working not seeking	Count	4	9	13
			% within Model	0.30%	0.40%	0.30%
		vol work not working not seeking	Count	1	3	4
			% within Model	0.10%	0.10%	0.10%
		retired	Count	76	103	179
			% within Model	4.90%	4.20%	4.50%
	Total		Count	1555	2447	4002
			% within Model	100.00%	100.00%	100.00%
Reliable Deterioration	Employment Status First	Employed	Count	107	139	246
			% within Model	35.00%	33.10%	33.90%
		Unemployed and Seeking Work	Count	109	132	241
			% within Model	35.60%	31.40%	33.20%
		ft students or part students not seeking work	Count	17	29	46
			% within Model	5.60%	6.90%	6.30%
		sick, disabled, incapacity ben, income support, employ & support allowance	Count	39	76	115
			% within Model	12.70%	18.10%	15.80%
		homemaker	Count	22	28	50
			% within Model	7.20%	6.70%	6.90%
		retired	Count	12	16	28
			% within Model	3.90%	3.80%	3.90%
	Total		Count	306	420	726
			% within Model	100.00%	100.00%	100.00%

Recovery				Model		Total
				Allocated	Progressive	
Total	Employment Status First	Employed	Count	2599	4519	7118
			% within Model	46.80%	46.60%	46.70%
		Unemployed and Seeking Work	Count	1480	2057	3537
			% within Model	26.60%	21.20%	23.20%
		ft students or part students not seeking work	Count	321	682	1003
			% within Model	5.80%	7.00%	6.60%
		sick, disabled, incapacity ben, income support, employ & support allowance	Count	517	1303	1820
			% within Model	9.30%	13.40%	11.90%
		homemaker	Count	326	511	837
			% within Model	5.90%	5.30%	5.50%
		no benefits not working not seeking	Count	4	32	36
			% within Model	0.10%	0.30%	0.20%
		vol work not working not seeking	Count	1	5	6
			% within Model	0.00%	0.10%	0.00%
		retired	Count	307	581	888
			% within Model	5.50%	6.00%	5.80%
	Total		Count	5555	9690	15245
			% within Model	100.00%	100.00%	100.00%

Appendix 50: Chi square for first employment status against model and recovery outcome for whole service

Chi-Square Tests

Recovery		Value	df	Asymp. Sig. (2-sided)
Recovered	Pearson Chi-Square	58.270[b]	7	0
	Likelihood Ratio	63.8	7	0
	Linear-by-Linear Association	7.697	1	0.006
	N of Valid Cases	6774		
Reliable Improvement	Pearson Chi-Square	28.512[c]	7	0
	Likelihood Ratio	32.263	7	0
	Linear-by-Linear Association	7.451	1	0.006
	N of Valid Cases	3743		
Non Recovered	Pearson Chi-Square	42.227[d]	7	0
	Likelihood Ratio	43.086	7	0
	Linear-by-Linear Association	2.35	1	0.125
	N of Valid Cases	4002		
Reliable Deterioration	Pearson Chi-Square	4.904[e]	5	0.428
	Likelihood Ratio	4.978	5	0.419
	Linear-by-Linear Association	0.857	1	0.355
	N of Valid Cases	726		
Total	Pearson Chi-Square	118.435[a]	7	0
	Likelihood Ratio	121.823	7	0
	Linear-by-Linear Association	14.268	1	0
	N of Valid Cases	15245		

Appendix 51: Cross tabulation of model and outcome for north cohort

			Model		Total
			Allocated	Progressive	
Recovery	Recovered	Count	1387	2571	3958
		% within Model	42.80%	46.90%	45.40%
	Reliable Improvement	Count	796	1294	2090
		% within Model	24.60%	23.60%	23.90%
	Non Recovered	Count	891	1360	2251
		% within Model	27.50%	24.80%	25.80%
	Reliable Deterioration	Count	168	260	428
		% within Model	5.20%	4.70%	4.90%
Total		Count	3242	5485	8727
		% within Model	100.00%	100.00%	100.00%

Appendix 52: Chi Square of model and outcome for north cohort

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	14.825[a]	3	0.002
Likelihood Ratio	14.829	3	0.002
Linear-by-Linear Association	13.32	1	0
N of Valid Cases	8727		

Appendix 53: Cross tabulation of model on recovery v non recovery for north cohort

			Model		Total
			Allocated	Progressive	
Recovery Non Recovery	Recovered	Count	1387	2571	3958
		% within Model	42.80%	46.90%	45.40%
	Non Recovered	Count	1855	2914	4769
		% within Model	57.20%	53.10%	54.60%
Total		Count	3242	5485	8727
		% within Model	100.00%	100.00%	100.00%

Appendix 54: Chi square of model on recovery v non recovery for north cohort

	Value	df	Asymp. Sig. (2-	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	13.760[a]	1	0		
Continuity Correction[b]	13.596	1	0		
Likelihood Ratio	13.783	1	0		
Fisher's Exact Test				0	0
Linear-by-Linear Association	13.759	1	0		
N of Valid Cases	8727				

Appendix 55: Cross tabulation of model on reliable improvement v non recovery for north cohort

			Model		Total
			Allocated	Progressive	
Reliable Improvement Non Recovery	Reliable Improvement	Count	796	1294	2090
		% within Model	42.90%	44.40%	43.80%
	Non Recovery	Count	1059	1620	2679
		% within Model	57.10%	55.60%	56.20%
Total		Count	1855	2914	4769
		% within Model	100.00%	100.00%	100.00%

Appendix 56: Chi square of model on reliable improvement v non recovery for north cohort

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1.029[a]	1	0.31		
Continuity Correction[b]	0.97	1	0.325		
Likelihood Ratio	1.03	1	0.31		
Fisher's Exact Test				0.323	0.162
Linear-by-Linear Association	1.029	1	0.31		
N of Valid Cases	4769				

Appendix 57: Cross tabulation of model on reliable deterioration v non recovery for north cohort

			Model		Total
			Allocated	Progressive	
RD V Non Recovery	Reliable Deterioration	Count	168	260	428
		% within Model	9.10%	8.90%	9.00%
	Non Recovery	Count	1687	2654	4341
		% within Model	90.90%	91.10%	91.00%
Total		Count	1855	2914	4769
		% within Model	100.00%	100.00%	100.00%

Appendix 58: Chi square of model on reliable deterioration v non recovery for north cohort

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.025[a]	1	0.874		
Continuity Correction[b]	0.011	1	0.916		
Likelihood Ratio	0.025	1	0.874		
Fisher's Exact Test				0.876	0.457
Linear-by-Linear Association	0.025	1	0.874		
N of Valid Cases	4769				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 166.48.

b. Computed only for a 2x2 table

Appendix 59: Cross tabulation on PHQ Moderate to Severe/Severe on model v outcomes for north cohort

PHQ Severity				Model		Total
				Allocated	Progressive	
Severe	Recovery	Recovered	Count	271	532	803
			% within Model	25.90%	30.60%	28.80%
		Reliable Improvement	Count	367	619	986
			% within Model	35.10%	35.60%	35.40%
		Non Recovered	Count	382	547	929
			% within Model	36.60%	31.40%	33.40%
		Reliable Deterioration	Count	25	42	67
			% within Model	2.40%	2.40%	2.40%
	Total		Count	1045	1740	2785
			% within Model	100.00%	100.00%	100.00%
Moderate to Severe	Recovery	Recovered	Count	404	804	1208
			% within Model	40.20%	46.00%	43.90%
		Reliable Improvement	Count	287	430	717
			% within Model	28.50%	24.60%	26.00%
		Non Recovered	Count	262	426	688
			% within Model	26.00%	24.40%	25.00%
		Reliable Deterioration	Count	53	88	141
			% within Model	5.30%	5.00%	5.10%
	Total		Count	1006	1748	2754
			% within Model	100.00%	100.00%	100.00%
Total	Recovery	Recovered	Count	675	1336	2011
			% within Model	32.90%	38.30%	36.30%
		Reliable Improvement	Count	654	1049	1703
			% within Model	31.90%	30.10%	30.70%
		Non Recovered	Count	644	973	1617
			% within Model	31.40%	27.90%	29.20%
		Reliable Deterioration	Count	78	130	208
			% within Model	3.80%	3.70%	3.80%
	Total		Count	2051	3488	5539
			% within Model	100.00%	100.00%	100.00%

Appendix 60: Chi square on PHQ Moderate to Severe/Severe on model v outcomes for north cohort

PHQ Severity		Value	df	Asymp. Sig. (2-sided)
Severe	Pearson Chi-Square	10.045[b]	3	0.018
	Likelihood Ratio	10.058	3	0.018
	Linear-by-Linear Association	8.687	1	0.003
	N of Valid Cases	2785		
Moderate to Severe	Pearson Chi-Square	9.529[c]	3	0.023
	Likelihood Ratio	9.541	3	0.023
	Linear-by-Linear Association	4.586	1	0.032
	N of Valid Cases	2754		
Total	Pearson Chi-Square	17.173[a]	3	0.001
	Likelihood Ratio	17.258	3	0.001
	Linear-by-Linear Association	13.128	1	0
	N of Valid Cases	5539		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 77.02.
b. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 25.14.
c. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 51.51.

Appendix 61: Cross tabulation on GAD Moderate/Severe on model v outcomes for north cohort

GAD Severity				Model		Total
				Allocated	Progressive	
Severe	Recovery	Recovered	Count	570	1133	1703
			% within Model	31.90%	37.00%	35.10%
		Reliable Improvement	Count	614	1003	1617
			% within Model	34.40%	32.70%	33.30%
		Non Recovered	Count	551	853	1404
			% within Model	30.90%	27.80%	28.90%
		Reliable Deterioration	Count	51	75	126
			% within Model	2.90%	2.40%	2.60%
	Total		Count	1786	3064	4850
			% within Model	100.00%	100.00%	100.00%
Moderate	Recovery	Recovered	Count	520	945	1465
			% within Model	52.80%	55.70%	54.60%
		Reliable Improvement	Count	151	260	411
			% within Model	15.30%	15.30%	15.30%
		Non Recovered	Count	246	373	619
			% within Model	25.00%	22.00%	23.10%
		Reliable Deterioration	Count	68	119	187
			% within Model	6.90%	7.00%	7.00%
	Total		Count	985	1697	2682
			% within Model	100.00%	100.00%	100.00%
Total	Recovery	Recovered	Count	1090	2078	3168
			% within Model	39.30%	43.60%	42.10%
		Reliable Improvement	Count	765	1263	2028
			% within Model	27.60%	26.50%	26.90%
		Non Recovered	Count	797	1226	2023
			% within Model	28.80%	25.80%	26.90%
		Reliable Deterioration	Count	119	194	313
			% within Model	4.30%	4.10%	4.20%
	Total		Count	2771	4761	7532
			% within Model	100.00%	100.00%	100.00%

Appendix 62: Chi square on GAD Moderate/Severe on model v outcomes for north cohort

GAD Severity		Value	df	Asymp. Sig. (2-sided)
Severe	Pearson Chi-Square	13.408[b]	3	0.004
	Likelihood Ratio	13.478	3	0.004
	Linear-by-Linear Association	11.976	1	0.001
	N of Valid Cases	4850		
Moderate	Pearson Chi-Square	3.388[c]	3	0.336
	Likelihood Ratio	3.367	3	0.338
	Linear-by-Linear Association	1.955	1	0.162
	N of Valid Cases	2682		
Total	Pearson Chi-Square	14.612[a]	3	0.002
	Likelihood Ratio	14.627	3	0.002
	Linear-by-Linear Association	12.401	1	0
	N of Valid Cases	7532		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 115.15.
b. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 46.40.
c. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 68.68.

Appendix 63: Cross tabulation for age against model and recovery outcome for north cohort

Recovery				Model		Total
				Allocated	Progressive	
Recovered	Age	16-17	Count	0	1	1
			% within Model	0.00%	0.00%	0.00%
		18-24	Count	87	310	397
			% within Model	6.30%	12.10%	10.00%
		25-34	Count	318	588	906
			% within Model	22.90%	22.90%	22.90%
		35-44	Count	316	590	906
			% within Model	22.80%	22.90%	22.90%
		45-54	Count	337	535	872
			% within Model	24.30%	20.80%	22.00%
		55-64	Count	207	360	567
			% within Model	14.90%	14.00%	14.30%
		65-74	Count	98	137	235
			% within Model	7.10%	5.30%	5.90%
		75-84	Count	18	40	58
			% within Model	1.30%	1.60%	1.50%
		85-94	Count	6	9	15
			% within Model	0.40%	0.40%	0.40%
		95+	Count	0	1	1
			% within Model	0.00%	0.00%	0.00%
	Total		Count	1387	2571	3958
			% within Model	100.00%	100.00%	100.00%
Reliable Improvement	Age	18-24	Count	49	156	205
			% within Model	6.20%	12.10%	9.80%
		25-34	Count	205	352	557
			% within Model	25.80%	27.20%	26.70%
		35-44	Count	192	270	462
			% within Model	24.10%	20.90%	22.10%
		45-54	Count	188	264	452
			% within Model	23.60%	20.40%	21.60%
		55-64	Count	128	201	329
			% within Model	16.10%	15.50%	15.70%
		65-74	Count	29	36	65
			% within Model	3.60%	2.80%	3.10%
		75-84	Count	4	13	17
			% within Model	0.50%	1.00%	0.80%
		85-94	Count	1	2	3
			% within Model	0.10%	0.20%	0.10%
	Total		Count	796	1294	2090
			% within Model	100.00%	100.00%	100.00%

Recovery				Model		Total
				Allocated	Progressive	
Non Recovered	Age	16-17	Count	0	1	1
			% within Model	0.00%	0.10%	0.00%
		18-24	Count	79	198	277
			% within Model	8.90%	14.60%	12.30%
		25-34	Count	234	333	567
			% within Model	26.30%	24.50%	25.20%
		35-44	Count	207	306	513
			% within Model	23.20%	22.50%	22.80%
		45-54	Count	196	287	483
			% within Model	22.00%	21.10%	21.50%
		55-64	Count	130	191	321
			% within Model	14.60%	14.00%	14.30%
		65-74	Count	38	32	70
			% within Model	4.30%	2.40%	3.10%
		75-84	Count	6	11	17
			% within Model	0.70%	0.80%	0.80%
		85-94	Count	1	1	2
			% within Model	0.10%	0.10%	0.10%
	Total		Count	891	1360	2251
			% within Model	100.00%	100.00%	100.00%
Reliable Deterioration	Age	18-24	Count	16	44	60
			% within Model	9.50%	16.90%	14.00%
		25-34	Count	41	52	93
			% within Model	24.40%	20.00%	21.70%
		35-44	Count	51	67	118
			% within Model	30.40%	25.80%	27.60%
		45-54	Count	30	55	85
			% within Model	17.90%	21.20%	19.90%
		55-64	Count	24	32	56
			% within Model	14.30%	12.30%	13.10%
		65-74	Count	6	10	16
			% within Model	3.60%	3.80%	3.70%
	Total		Count	168	260	428
			% within Model	100.00%	100.00%	100.00%

Recovery				Model		Total
				Allocated	Progressive	
Total	Age	16-17	Count	0	2	2
			% within Model	0.00%	0.00%	0.00%
		18-24	Count	231	708	939
			% within Model	7.10%	12.90%	10.80%
		25-34	Count	798	1325	2123
			% within Model	24.60%	24.20%	24.30%
		35-44	Count	766	1233	1999
			% within Model	23.60%	22.50%	22.90%
		45-54	Count	751	1141	1892
			% within Model	23.20%	20.80%	21.70%
		55-64	Count	489	784	1273
			% within Model	15.10%	14.30%	14.60%
		65-74	Count	171	215	386
			% within Model	5.30%	3.90%	4.40%
		75-84	Count	28	64	92
			% within Model	0.90%	1.20%	1.10%
		85-94	Count	8	12	20
			% within Model	0.20%	0.20%	0.20%
		95+	Count	0	1	1
			% within Model	0.00%	0.00%	0.00%
	Total		Count	3242	5485	8727
			% within Model	100.00%	100.00%	100.00%

Appendix 64: Chi square for age against model and recovery outcome for north cohort

Recovery		Value	df	Asymp. Sig. (2-sided)
Recovered	Pearson Chi-Square	41.811[b]	9	0
	Likelihood Ratio	44.66	9	0
	Linear-by-Linear Association	18.72	1	0
	N of Valid Cases	3958		
Reliable Improvement	Pearson Chi-Square	25.422[c]	7	0.001
	Likelihood Ratio	26.674	7	0
	Linear-by-Linear Association	8.346	1	0.004
	N of Valid Cases	2090		
Non Recovered	Pearson Chi-Square	22.495[d]	8	0.004
	Likelihood Ratio	23.292	8	0.003
	Linear-by-Linear Association	8.21	1	0.004
	N of Valid Cases	2251		
Reliable Deterioration	Pearson Chi-Square	6.560[e]	5	0.255
	Likelihood Ratio	6.758	5	0.239
	Linear-by-Linear Association	0.582	1	0.446
	N of Valid Cases	428		
Total	Pearson Chi-Square	82.865[a]	9	0
	Likelihood Ratio	87.701	9	0
	Linear-by-Linear Association	32.376	1	0
	N of Valid Cases	8727		

Appendix 65: Cross tabulation for gender against model and recovery outcome for north cohort

Recovery				Model		Total
				Allocated	Progressive	
Recovered	Gender	Male	Count	505	983	1488
			% within Model	36.40%	38.20%	37.60%
		Female	Count	882	1588	2470
			% within Model	63.60%	61.80%	62.40%
	Total		Count	1387	2571	3958
			% within Model	100.00%	100.00%	100.00%
Reliable Improvement	Gender	Male	Count	303	485	788
			% within Model	38.10%	37.50%	37.70%
		Female	Count	493	808	1301
			% within Model	61.90%	62.50%	62.30%
	Total		Count	796	1293	2089
			% within Model	100.00%	100.00%	100.00%
Non Recovered	Gender	Male	Count	346	513	859
			% within Model	38.90%	37.70%	38.20%
		Female	Count	544	847	1391
			% within Model	61.10%	62.30%	61.80%
	Total		Count	890	1360	2250
			% within Model	100.00%	100.00%	100.00%
Reliable Deterioration	Gender	Male	Count	58	91	149
			% within Model	34.50%	35.00%	34.80%
		Female	Count	110	169	279
			% within Model	65.50%	65.00%	65.20%
	Total		Count	168	260	428
			% within Model	100.00%	100.00%	100.00%
Total	Gender	Male	Count	1212	2072	3284
			% within Model	37.40%	37.80%	37.60%
		Female	Count	2029	3412	5441
			% within Model	62.60%	62.20%	62.40%
	Total		Count	3241	5484	8725
			% within Model	100.00%	100.00%	100.00%

Appendix 66: Chi square for gender against model and recovery outcome for north cohort

Recovery		Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Recovered	Pearson Chi-Square	1.279[c]	1	0.258		
	Continuity Correction[b]	1.202	1	0.273		
	Likelihood Ratio	1.281	1	0.258		
	Fisher's Exact Test				0.271	0.136
	Linear-by-Linear Association	1.278	1	0.258		
	N of Valid Cases	3958				
Reliable Improvement	Pearson Chi-Square	.065[d]	1	0.799		
	Continuity Correction[b]	0.043	1	0.835		
	Likelihood Ratio	0.065	1	0.799		
	Fisher's Exact Test				0.816	0.417
	Linear-by-Linear Association	0.065	1	0.799		
	N of Valid Cases	2089				
Non Recovered	Pearson Chi-Square	.304[e]	1	0.581		
	Continuity Correction[b]	0.257	1	0.612		
	Likelihood Ratio	0.304	1	0.581		
	Fisher's Exact Test				0.595	0.306
	Linear-by-Linear Association	0.304	1	0.581		
	N of Valid Cases	2250				
Reliable Deterioration	Pearson Chi-Square	.010[f]	1	0.92		
	Continuity Correction[b]	0	1	1		
	Likelihood Ratio	0.01	1	0.92		
	Fisher's Exact Test				1	0.502
	Linear-by-Linear Association	0.01	1	0.92		
	N of Valid Cases	428				
Total	Pearson Chi-Square	.130[a]	1	0.719		
	Continuity Correction[b]	0.114	1	0.736		
	Likelihood Ratio	0.13	1	0.719		
	Fisher's Exact Test				0.732	0.368
	Linear-by-Linear Association	0.13	1	0.719		
	N of Valid Cases	8725				

Appendix 67: Cross tabulation for disability against model and recovery outcome for north cohort

Recovery				Model		Total
				Allocated	Progressive	
Recovered	Disability	Yes	Count	43	235	278
			% within Model	3.10%	9.10%	7.00%
		No	Count	1344	2336	3680
			% within Model	96.90%	90.90%	93.00%
	Total		Count	1387	2571	3958
			% within Model	100.00%	100.00%	100.00%
Reliable Improvement	Disability	Yes	Count	25	171	196
			% within Model	3.10%	13.20%	9.40%
		No	Count	771	1123	1894
			% within Model	96.90%	86.80%	90.60%
	Total		Count	796	1294	2090
			% within Model	100.00%	100.00%	100.00%
Non Recovered	Disability	Yes	Count	41	167	208
			% within Model	4.60%	12.30%	9.20%
		No	Count	850	1193	2043
			% within Model	95.40%	87.70%	90.80%
	Total		Count	891	1360	2251
			% within Model	100.00%	100.00%	100.00%
Reliable Deterioration	Disability	Yes	Count	4	26	30
			% within Model	2.40%	10.00%	7.00%
		No	Count	164	234	398
			% within Model	97.60%	90.00%	93.00%
	Total		Count	168	260	428
			% within Model	100.00%	100.00%	100.00%
Total	Disability	Yes	Count	113	599	712
			% within Model	3.50%	10.90%	8.20%
		No	Count	3129	4886	8015
			% within Model	96.50%	89.10%	91.80%
	Total		Count	3242	5485	8727
			% within Model	100.00%	100.00%	100.00%

Appendix 68: Chi square for disability against model and recovery outcome for north cohort

Recovery		Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Recovered	Pearson Chi-Square	50.334[c]	1	0		
	Continuity Correction[b]	49.414	1	0		
	Likelihood Ratio	56.982	1	0		
	Fisher's Exact Test				0	0
	Linear-by-Linear Association	50.322	1	0		
	N of Valid Cases	3958				
Reliable Improvement	Pearson Chi-Square	58.854[d]	1	0		
	Continuity Correction[b]	57.674	1	0		
	Likelihood Ratio	68.075	1	0		
	Fisher's Exact Test				0	0
	Linear-by-Linear Association	58.825	1	0		
	N of Valid Cases	2090				
Non Recovered	Pearson Chi-Square	37.839[e]	1	0		
	Continuity Correction[b]	36.929	1	0		
	Likelihood Ratio	41.28	1	0		
	Fisher's Exact Test				0	0
	Linear-by-Linear Association	37.822	1	0		
	N of Valid Cases	2251				
Reliable Deterioration	Pearson Chi-Square	9.089[f]	1	0.003		
	Continuity Correction[b]	7.958	1	0.005		
	Likelihood Ratio	10.473	1	0.001		
	Fisher's Exact Test				0.003	0.001
	Linear-by-Linear Association	9.068	1	0.003		
	N of Valid Cases	428				
Total	Pearson Chi-Square	150.333[a]	1	0		
	Continuity Correction[b]	149.343	1	0		
	Likelihood Ratio	169.303	1	0		
	Fisher's Exact Test				0	0
	Linear-by-Linear Association	150.316	1	0		
	N of Valid Cases	8727				

Appendix 69: Cross tabulation for first employment status against model and recovery outcome for north cohort

Recovery				Model		Total
				Allocated	Progressive	
Recovered	Employment Status First	Employed	Count	786	1388	2174
			% within Model	57.30%	54.20%	55.30%
		Unemployed and Seeking Work	Count	240	328	568
			% within Model	17.50%	12.80%	14.40%
		ft students or part students not seeking work	Count	111	235	346
			% within Model	8.10%	9.20%	8.80%
		sick, disabled, incapacity ben, income support, employ & support allowance	Count	80	285	365
			% within Model	5.80%	11.10%	9.30%
		homemaker	Count	56	112	168
			% within Model	4.10%	4.40%	4.30%
		no benefits not working not seeking	Count	0	12	12
			% within Model	0.00%	0.50%	0.30%
		vol work not working not seeking	Count	0	1	1
			% within Model	0.00%	0.00%	0.00%
		retired	Count	98	201	299
			% within Model	7.10%	7.80%	7.60%
	Total		Count	1371	2562	3933
			% within Model	100.00%	100.00%	100.00%
Reliable Improvement	Employment Status First	Employed	Count	331	507	838
			% within Model	41.90%	39.30%	40.30%
		Unemployed and Seeking Work	Count	229	293	522
			% within Model	29.00%	22.70%	25.10%
		ft students or part students not seeking work	Count	47	105	152
			% within Model	5.90%	8.10%	7.30%
		sick, disabled, incapacity ben, income support, employ & support allowance	Count	110	267	377
			% within Model	13.90%	20.70%	18.10%
		homemaker	Count	40	54	94
			% within Model	5.10%	4.20%	4.50%
		no benefits not working not seeking	Count	0	7	7
			% within Model	0.00%	0.50%	0.30%
		vol work not working not seeking	Count	0	1	1
			% within Model	0.00%	0.10%	0.00%
		retired	Count	33	57	90
			% within Model	4.20%	4.40%	4.30%
	Total		Count	790	1291	2081
			% within Model	100.00%	100.00%	100.00%

Recovery				Model		Total
				Allocated	Progressive	
Non Recovered	Employment Status First	Employed	Count	290	455	745
			% within Model	32.90%	33.60%	33.30%
		Unemployed and Seeking Work	Count	301	340	641
			% within Model	34.10%	25.10%	28.60%
		ft students or part students not seeking work	Count	61	124	185
			% within Model	6.90%	9.10%	8.30%
		sick, disabled, incapacity ben, income support, employ & support allowance	Count	133	307	440
			% within Model	15.10%	22.60%	19.70%
		homemaker	Count	50	72	122
			% within Model	5.70%	5.30%	5.50%
		no benefits not working not seeking	Count	2	8	10
			% within Model	0.20%	0.60%	0.40%
		vol work not working not seeking	Count	1	3	4
			% within Model	0.10%	0.20%	0.20%
		retired	Count	44	47	91
			% within Model	5.00%	3.50%	4.10%
	Total		Count	882	1356	2238
			% within Model	100.00%	100.00%	100.00%
Reliable Deterioration	Employment Status First	Employed	Count	60	90	150
			% within Model	36.40%	34.90%	35.50%
		Unemployed and Seeking Work	Count	52	69	121
			% within Model	31.50%	26.70%	28.60%
		ft students or part students not seeking work	Count	14	25	39
			% within Model	8.50%	9.70%	9.20%
		sick, disabled, incapacity ben, income support, employ & support allowance	Count	24	53	77
			% within Model	14.50%	20.50%	18.20%
		homemaker	Count	9	13	22
			% within Model	5.50%	5.00%	5.20%
		retired	Count	6	8	14
			% within Model	3.60%	3.10%	3.30%
	Total		Count	165	258	423
			% within Model	100.00%	100.00%	100.00%

Recovery				Model		Total
				Allocated	Progressive	
Total	Employment Status First	Employed	Count	1467	2440	3907
			% within Model	45.70%	44.60%	45.00%
		Unemployed and Seeking Work	Count	822	1030	1852
			% within Model	25.60%	18.80%	21.30%
		ft students or part students not seeking work	Count	233	489	722
			% within Model	7.30%	8.90%	8.30%
		sick, disabled, incapacity ben, income support, employ & support allowance	Count	347	912	1259
			% within Model	10.80%	16.70%	14.50%
		homemaker	Count	155	251	406
			% within Model	4.80%	4.60%	4.70%
		no benefits not working not seeking	Count	2	27	29
			% within Model	0.10%	0.50%	0.30%
		vol work not working not seeking	Count	1	5	6
			% within Model	0.00%	0.10%	0.10%
		retired	Count	181	313	494
			% within Model	5.60%	5.70%	5.70%
	Total		Count	3208	5467	8675
			% within Model	100.00%	100.00%	100.00%

Appendix 70: Chi square for first employment status against model and recovery outcome for north cohort

Recovery		Value	df	Asymp. Sig. (2-sided)
Recovered	Pearson Chi-Square	51.080[b]	7	0
	Likelihood Ratio	57.028	7	0
	Linear-by-Linear Association	10.499	1	0.001
	N of Valid Cases	3933		
Reliable Improvement	Pearson Chi-Square	29.929[c]	7	0
	Likelihood Ratio	33.079	7	0
	Linear-by-Linear Association	6.248	1	0.012
	N of Valid Cases	2081		
Non Recovered	Pearson Chi-Square	39.213[d]	7	0
	Likelihood Ratio	39.671	7	0
	Linear-by-Linear Association	1.299	1	0.254
	N of Valid Cases	2238		
Reliable Deterioration	Pearson Chi-Square	3.131[e]	5	0.68
	Likelihood Ratio	3.181	5	0.672
	Linear-by-Linear Association	0.399	1	0.528
	N of Valid Cases	423		
Total	Pearson Chi-Square	111.500[a]	7	0
	Likelihood Ratio	115.996	7	0
	Linear-by-Linear Association	15.47	1	0
	N of Valid Cases	8675		

Appendix 71: Cross tabulation of model and outcome for south cohort

			Model		Total
			Allocated	Progressive	
Recovery	Recovered	Count	1764	1045	2809
		% within Model	40.10%	49.70%	43.20%
	Reliable Improvement	Count	1153	493	1646
		% within Model	26.20%	23.50%	25.30%
	Non Recovered	Count	1253	488	1741
		% within Model	28.50%	23.20%	26.80%
	Reliable Deterioration	Count	228	76	304
		% within Model	5.20%	3.60%	4.70%
Total		Count	4398	2102	6500
		% within Model	100.00%	100.00%	100.00%

Appendix 72: Chi square of model and outcome for south cohort

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	56.904[a]	3	0
Likelihood Ratio	56.976	3	0
Linear-by-Linear Association	52.331	1	0
N of Valid Cases	6500		

Appendix 73: Cross tabulation of model on recovery v non recovery for south cohort

			Model		Total
			Allocated	Progressive	
Recovery Non Recovery	Recovered	Count	1764	1045	2809
		% within Model	40.10%	49.70%	43.20%
	Non Recovered	Count	2634	1057	3691
		% within Model	59.90%	50.30%	56.80%
Total		Count	4398	2102	6500
		% within Model	100.00%	100.00%	100.00%

Appendix 74: Chi square of model on recovery v non recovery for south cohort

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	53.473[a]	1	0		
Continuity Correction[b]	53.083	1	0		
Likelihood Ratio	53.259	1	0		
Fisher's Exact Test				0	0
Linear-by-Linear Association	53.465	1	0		
N of Valid Cases	6500				

Appendix 75: Cross tabulation of model on reliable improvement v non recovery for south cohort

			Model		Total
			Allocated	Progressive	
Reliable Improvement Non Recovery	Reliable Improvement	Count	1153	493	1646
		% within Model	43.80%	46.60%	44.60%
	Non Recovery	Count	1481	564	2045
		% within Model	56.20%	53.40%	55.40%
Total		Count	2634	1057	3691
		% within Model	100.00%	100.00%	100.00%

Appendix 76: Chi square of model on reliable improvement v non recovery for south cohort

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	2.511[a]	1	0.113		
Continuity Correction[b]	2.396	1	0.122		
Likelihood Ratio	2.507	1	0.113		
Fisher's Exact Test				0.115	0.061
Linear-by-Linear Association	2.51	1	0.113		
N of Valid Cases	3691				

Appendix 77: Cross tabulation of model on reliable deterioration v non recovery for south cohort

			Model		Total
			Allocated	Progressive	
RD V Non Recovery	Reliable Deterioration	Count	228	76	304
		% within Model	8.70%	7.20%	8.20%
	Non Recovery	Count	2406	981	3387
		% within Model	91.30%	92.80%	91.80%
Total		Count	2634	1057	3691
		% within Model	100.00%	100.00%	100.00%

Appendix 78: Chi square of model on reliable deterioration v non recovery for south cohort

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	2.145[a]	1	0.143		
Continuity Correction[b]	1.955	1	0.162		
Likelihood Ratio	2.198	1	0.138		
Fisher's Exact Test				0.146	0.08
Linear-by-Linear Association	2.144	1	0.143		
N of Valid Cases	3691				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 87.06.

b. Computed only for a 2x2 table

Appendix 79: Cross tabulation on PHQ Moderate to Severe/Severe on model v outcomes south cohort

PHQ Severity				Model		Total
				Allocated	Progressive	
Severe	Recovery	Recovered	Count	349	223	572
			% within Model	23.70%	32.10%	26.40%
		Reliable Improvement	Count	544	246	790
			% within Model	36.90%	35.40%	36.40%
		Non Recovered	Count	546	208	754
			% within Model	37.00%	29.90%	34.80%
		Reliable Deterioration	Count	35	18	53
			% within Model	2.40%	2.60%	2.40%
	Total		Count	1474	695	2169
			% within Model	100.00%	100.00%	100.00%
Moderate to Severe	Recovery	Recovered	Count	485	317	802
			% within Model	37.20%	47.80%	40.80%
		Reliable Improvement	Count	385	163	548
			% within Model	29.50%	24.60%	27.90%
		Non Recovered	Count	354	156	510
			% within Model	27.10%	23.50%	25.90%
		Reliable Deterioration	Count	80	27	107
			% within Model	6.10%	4.10%	5.40%
	Total		Count	1304	663	1967
			% within Model	100.00%	100.00%	100.00%
Total	Recovery	Recovered	Count	834	540	1374
			% within Model	30.00%	39.80%	33.20%
		Reliable Improvement	Count	929	409	1338
			% within Model	33.40%	30.10%	32.40%
		Non Recovered	Count	900	364	1264
			% within Model	32.40%	26.80%	30.60%
		Reliable Deterioration	Count	115	45	160
			% within Model	4.10%	3.30%	3.90%
	Total		Count	2778	1358	4136
			% within Model	100.00%	100.00%	100.00%

Appendix 80: Chi square on PHQ Moderate to Severe/Severe on model v outcomes for south cohort

PHQ Severity		Value	df	Asymp. Sig. (2-sided)
Severe	Pearson Chi-Square	19.927[b]	3	0
	Likelihood Ratio	19.697	3	0
	Linear-by-Linear Association	15.552	1	0
	N of Valid Cases	2169		
Moderate to Severe	Pearson Chi-Square	21.663[c]	3	0
	Likelihood Ratio	21.651	3	0
	Linear-by-Linear Association	16.775	1	0
	N of Valid Cases	1967		
Total	Pearson Chi-Square	40.122[a]	3	0
	Likelihood Ratio	39.66	3	0
	Linear-by-Linear Association	33.313	1	0
	N of Valid Cases	4136		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 52.53.
b. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 16.98.
c. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 36.07.

Appendix 81: Cross tabulation on GAD Moderate/Severe on model v outcomes for south cohort

GAD Severity				Model		Total
				Allocated	Progressive	
Severe	Recovery	Recovered	Count	751	463	1214
			% within Model	30.40%	39.00%	33.20%
		Reliable Improvement	Count	861	397	1258
			% within Model	34.80%	33.40%	34.40%
		Non Recovered	Count	791	309	1100
			% within Model	32.00%	26.00%	30.10%
		Reliable Deterioration	Count	68	19	87
			% within Model	2.80%	1.60%	2.40%
	Total		Count	2471	1188	3659
			% within Model	100.00%	100.00%	100.00%
Moderate	Recovery	Recovered	Count	651	391	1042
			% within Model	48.50%	61.60%	52.70%
		Reliable Improvement	Count	252	86	338
			% within Model	18.80%	13.50%	17.10%
		Non Recovered	Count	338	122	460
			% within Model	25.20%	19.20%	23.30%
		Reliable Deterioration	Count	101	36	137
			% within Model	7.50%	5.70%	6.90%
	Total		Count	1342	635	1977
			% within Model	100.00%	100.00%	100.00%
Total	Recovery	Recovered	Count	1402	854	2256
			% within Model	36.80%	46.80%	40.00%
		Reliable Improvement	Count	1113	483	1596
			% within Model	29.20%	26.50%	28.30%
		Non Recovered	Count	1129	431	1560
			% within Model	29.60%	23.60%	27.70%
		Reliable Deterioration	Count	169	55	224
			% within Model	4.40%	3.00%	4.00%
	Total		Count	3813	1823	5636
			% within Model	100.00%	100.00%	100.00%

Appendix 82: Chi square on GAD Moderate/Severe on model v outcomes for south cohort

GAD Severity		Value	df	Asymp. Sig. (2-sided)
Severe	Pearson Chi-Square	32.372[b]	3	0
	Likelihood Ratio	32.474	3	0
	Linear-by-Linear Association	31.45	1	0
	N of Valid Cases	3659		
Moderate	Pearson Chi-Square	29.624[c]	3	0
	Likelihood Ratio	29.859	3	0
	Linear-by-Linear Association	22.024	1	0
	N of Valid Cases	1977		
Total	Pearson Chi-Square	56.530[a]	3	0
	Likelihood Ratio	56.505	3	0
	Linear-by-Linear Association	52.664	1	0
	N of Valid Cases	5636		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 72.45.
b. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 28.25.
c. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 44.00.

Appendix 83: Cross tabulation for age against model and recovery outcome for south cohort

Recovery				Model		Total
				Allocated	Progressive	
Recovered	Age	16-17	Count	0	1	1
			% within Model	0.00%	0.10%	0.00%
		18-24	Count	124	98	222
			% within Model	7.00%	9.40%	7.90%
		25-34	Count	363	250	613
			% within Model	20.60%	23.90%	21.80%
		35-44	Count	441	257	698
			% within Model	25.00%	24.60%	24.80%
		45-54	Count	417	236	653
			% within Model	23.60%	22.60%	23.20%
		55-64	Count	280	130	410
			% within Model	15.90%	12.40%	14.60%
		65-74	Count	127	61	188
			% within Model	7.20%	5.80%	6.70%
		75-84	Count	11	10	21
			% within Model	0.60%	1.00%	0.70%
		85-94	Count	1	2	3
			% within Model	0.10%	0.20%	0.10%
	Total		Count	1764	1045	2809
			% within Model	100.00%	100.00%	100.00%
Reliable Improvement	Age	18-24	Count	80	61	141
			% within Model	6.90%	12.40%	8.60%
		25-34	Count	296	126	422
			% within Model	25.70%	25.60%	25.60%
		35-44	Count	257	110	367
			% within Model	22.30%	22.30%	22.30%
		45-54	Count	259	105	364
			% within Model	22.50%	21.30%	22.10%
		55-64	Count	209	73	282
			% within Model	18.10%	14.80%	17.10%
		65-74	Count	42	16	58
			% within Model	3.60%	3.20%	3.50%
		75-84	Count	10	1	11
			% within Model	0.90%	0.20%	0.70%
		85-94	Count	0	1	1
			% within Model	0.00%	0.20%	0.10%
	Total		Count	1153	493	1646
			% within Model	100.00%	100.00%	100.00%

Recovery				Model		Total
				Allocated	Progressive	
Non Recovered	Age	18-24	Count	110	81	191
			% within Model	8.80%	16.60%	11.00%
		25-34	Count	293	91	384
			% within Model	23.40%	18.60%	22.10%
		35-44	Count	272	122	394
			% within Model	21.70%	25.00%	22.60%
		45-54	Count	302	110	412
			% within Model	24.10%	22.50%	23.70%
		55-64	Count	229	65	294
			% within Model	18.30%	13.30%	16.90%
		65-74	Count	40	16	56
			% within Model	3.20%	3.30%	3.20%
		75-84	Count	5	3	8
			% within Model	0.40%	0.60%	0.50%
		85-94	Count	2	0	2
			% within Model	0.20%	0.00%	0.10%
	Total		Count	1253	488	1741
			% within Model	100.00%	100.00%	100.00%
Reliable Deterioration	Age	18-24	Count	16	9	25
			% within Model	7.00%	11.80%	8.20%
		25-34	Count	52	20	72
			% within Model	22.80%	26.30%	23.70%
		35-44	Count	56	16	72
			% within Model	24.60%	21.10%	23.70%
		45-54	Count	51	16	67
			% within Model	22.40%	21.10%	22.00%
		55-64	Count	38	12	50
			% within Model	16.70%	15.80%	16.40%
		65-74	Count	13	3	16
			% within Model	5.70%	3.90%	5.30%
		75-84	Count	2	0	2
			% within Model	0.90%	0.00%	0.70%
	Total		Count	228	76	304
			% within Model	100.00%	100.00%	100.00%

Recovery				Model		Total
				Allocated	Progressive	
Total	Age	16-17	Count	0	1	1
			% within Model	0.00%	0.00%	0.00%
		18-24	Count	330	249	579
			% within Model	7.50%	11.80%	8.90%
		25-34	Count	1004	487	1491
			% within Model	22.80%	23.20%	22.90%
		35-44	Count	1026	505	1531
			% within Model	23.30%	24.00%	23.60%
		45-54	Count	1029	467	1496
			% within Model	23.40%	22.20%	23.00%
		55-64	Count	756	280	1036
			% within Model	17.20%	13.30%	15.90%
		65-74	Count	222	96	318
			% within Model	5.00%	4.60%	4.90%
		75-84	Count	28	14	42
			% within Model	0.60%	0.70%	0.60%
		85-94	Count	3	3	6
			% within Model	0.10%	0.10%	0.10%
	Total		Count	4398	2102	6500
			% within Model	100.00%	100.00%	100.00%

Appendix 84: Chi square for age against model and recovery outcome for south cohort

Recovery		Value	df	Asymp. Sig. (2-sided)
Recovered	Pearson Chi-Square	19.199[b]	8	0.014
	Likelihood Ratio	19.434	8	0.013
	Linear-by-Linear Association	11.27	1	0.001
	N of Valid Cases	2809		
Reliable Improvement	Pearson Chi-Square	19.117[c]	7	0.008
	Likelihood Ratio	19.036	7	0.008
	Linear-by-Linear Association	8.443	1	0.004
	N of Valid Cases	1646		
Non Recovered	Pearson Chi-Square	31.442[d]	7	0
	Likelihood Ratio	30.779	7	0
	Linear-by-Linear Association	8.993	1	0.003
	N of Valid Cases	1741		
Reliable Deterioration	Pearson Chi-Square	3.277[e]	6	0.773
	Likelihood Ratio	3.66	6	0.723
	Linear-by-Linear Association	1.856	1	0.173
	N of Valid Cases	304		
Total	Pearson Chi-Square	48.328[a]	8	0
	Likelihood Ratio	47.507	8	0
	Linear-by-Linear Association	26.838	1	0
	N of Valid Cases	6500		

Appendix 85: Cross tabulation for gender against model and recovery outcome for south cohort

Recovery				Model		Total
				Allocated	Progressive	
Recovered	Gender	Male	Count	663	351	1014
			% within Model	37.60%	33.60%	36.10%
		Female	Count	1100	694	1794
			% within Model	62.40%	66.40%	63.90%
	Total		Count	1763	1045	2808
			% within Model	100.00%	100.00%	100.00%
Reliable Improvement	Gender	Male	Count	420	166	586
			% within Model	36.40%	33.70%	35.60%
		Female	Count	733	327	1060
			% within Model	63.60%	66.30%	64.40%
	Total		Count	1153	493	1646
			% within Model	100.00%	100.00%	100.00%
Non Recovered	Gender	Male	Count	478	185	663
			% within Model	38.20%	37.90%	38.10%
		Female	Count	774	303	1077
			% within Model	61.80%	62.10%	61.90%
	Total		Count	1252	488	1740
			% within Model	100.00%	100.00%	100.00%
Reliable Deterioration	Gender	Male	Count	80	24	104
			% within Model	35.10%	31.60%	34.20%
		Female	Count	148	52	200
			% within Model	64.90%	68.40%	65.80%
	Total		Count	228	76	304
			% within Model	100.00%	100.00%	100.00%
Total	Gender	Male	Count	1641	726	2367
			% within Model	37.30%	34.50%	36.40%
		Female	Count	2755	1376	4131
			% within Model	62.70%	65.50%	63.60%
	Total		Count	4396	2102	6498
			% within Model	100.00%	100.00%	100.00%

Appendix 86: Chi square for gender against model and recovery outcome for south cohort

Recovery		Value	df	Asymp. Sig. (2-	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Recovered	Pearson Chi-Square	4.591[c]	1	0.032		
	Continuity Correction[b]	4.418	1	0.036		
	Likelihood Ratio	4.612	1	0.032		
	Fisher's Exact Test				0.035	0.018
	Linear-by-Linear Association	4.589	1	0.032		
	N of Valid Cases	2808				
Reliable Improvement	Pearson Chi-Square	1.144[d]	1	0.285		
	Continuity Correction[b]	1.027	1	0.311		
	Likelihood Ratio	1.149	1	0.284		
	Fisher's Exact Test				0.312	0.155
	Linear-by-Linear Association	1.143	1	0.285		
	N of Valid Cases	1646				
Non Recovered	Pearson Chi-Square	.011[e]	1	0.917		
	Continuity Correction[b]	0.002	1	0.961		
	Likelihood Ratio	0.011	1	0.917		
	Fisher's Exact Test				0.956	0.481
	Linear-by-Linear Association	0.011	1	0.917		
	N of Valid Cases	1740				
Reliable Deterioration	Pearson Chi-Square	.312[f]	1	0.577		
	Continuity Correction[b]	0.175	1	0.675		
	Likelihood Ratio	0.315	1	0.575		
	Fisher's Exact Test				0.676	0.34
	Linear-by-Linear Association	0.311	1	0.577		
	N of Valid Cases	304				
Total	Pearson Chi-Square	4.783[a]	1	0.029		
	Continuity Correction[b]	4.663	1	0.031		
	Likelihood Ratio	4.803	1	0.028		
	Fisher's Exact Test				0.03	0.015
	Linear-by-Linear Association	4.782	1	0.029		
	N of Valid Cases	6498				

Appendix 87: Cross tabulation for disability against model and recovery outcome for south cohort

Recovery				Model		Total
				Allocated	Progressive	
Recovered	Disability	Yes	Count	87	77	164
			% within Model	4.90%	7.40%	5.80%
		No	Count	1677	968	2645
			% within Model	95.10%	92.60%	94.20%
	Total		Count	1764	1045	2809
			% within Model	100.00%	100.00%	100.00%
Reliable Improvement	Disability	Yes	Count	87	46	133
			% within Model	7.50%	9.30%	8.10%
		No	Count	1066	447	1513
			% within Model	92.50%	90.70%	91.90%
	Total		Count	1153	493	1646
			% within Model	100.00%	100.00%	100.00%
Non Recovered	Disability	Yes	Count	106	61	167
			% within Model	8.50%	12.50%	9.60%
		No	Count	1147	427	1574
			% within Model	91.50%	87.50%	90.40%
	Total		Count	1253	488	1741
			% within Model	100.00%	100.00%	100.00%
Reliable Deterioration	Disability	Yes	Count	24	11	35
			% within Model	10.50%	14.50%	11.50%
		No	Count	204	65	269
			% within Model	89.50%	85.50%	88.50%
	Total		Count	228	76	304
			% within Model	100.00%	100.00%	100.00%
Total	Disability	Yes	Count	304	195	499
			% within Model	6.90%	9.30%	7.70%
		No	Count	4094	1907	6001
			% within Model	93.10%	90.70%	92.30%
	Total		Count	4398	2102	6500
			% within Model	100.00%	100.00%	100.00%

Appendix 88: Chi square for disability against model and recovery outcome for south cohort

Recovery		Value	df	Asymp. Sig. (2-	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Recovered	Pearson Chi-Square	7.086[c]	1	0.008		
	Continuity Correction[b]	6.65	1	0.01		
	Likelihood Ratio	6.902	1	0.009		
	Fisher's Exact Test				0.01	0.005
	Linear-by-Linear Association	7.084	1	0.008		
	N of Valid Cases	2809				
Reliable Improvement	Pearson Chi-Square	1.482[d]	1	0.224		
	Continuity Correction[b]	1.251	1	0.263		
	Likelihood Ratio	1.446	1	0.229		
	Fisher's Exact Test				0.236	0.132
	Linear-by-Linear Association	1.481	1	0.224		
	N of Valid Cases	1646				
Non Recovered	Pearson Chi-Square	6.611[e]	1	0.01		
	Continuity Correction[b]	6.153	1	0.013		
	Likelihood Ratio	6.307	1	0.012		
	Fisher's Exact Test				0.014	0.007
	Linear-by-Linear Association	6.607	1	0.01		
	N of Valid Cases	1741				
Reliable Deterioration	Pearson Chi-Square	.872[f]	1	0.35		
	Continuity Correction[b]	0.527	1	0.468		
	Likelihood Ratio	0.834	1	0.361		
	Fisher's Exact Test				0.406	0.23
	Linear-by-Linear Association	0.869	1	0.351		
	N of Valid Cases	304				
Total	Pearson Chi-Square	11.220[a]	1	0.001		
	Continuity Correction[b]	10.889	1	0.001		
	Likelihood Ratio	10.898	1	0.001		
	Fisher's Exact Test				0.001	0.001
	Linear-by-Linear Association	11.219	1	0.001		
	N of Valid Cases	6500				

Appendix 89: Cross tabulation for first employment status against model and recovery outcome for south cohort

Recovery				Model		Total
				Allocated	Progressive	
Recovered	Employment Status First	Employed	Count	1089	663	1752
			% within Model	62.00%	63.60%	62.60%
		Unemployed and Seeking Work	Count	311	128	439
			% within Model	17.70%	12.30%	15.70%
		ft students or part students not seeking work	Count	69	43	112
			% within Model	3.90%	4.10%	4.00%
		sick, disabled, incapacity ben, income support, employ & support allowance	Count	72	64	136
			% within Model	4.10%	6.10%	4.90%
		homemaker	Count	91	64	155
			% within Model	5.20%	6.10%	5.50%
		no benefits not working not seeking	Count	2	0	2
			% within Model	0.10%	0.00%	0.10%
		retired	Count	123	80	203
			% within Model	7.00%	7.70%	7.30%
	Total		Count	1757	1042	2799
			% within Model	100.00%	100.00%	100.00%
Reliable Improvement	Employment Status First	Employed	Count	500	208	708
			% within Model	43.60%	42.40%	43.20%
		Unemployed and Seeking Work	Count	350	143	493
			% within Model	30.50%	29.20%	30.10%
		ft students or part students not seeking work	Count	51	24	75
			% within Model	4.40%	4.90%	4.60%
		sick, disabled, incapacity ben, income support, employ & support allowance	Count	112	53	165
			% within Model	9.80%	10.80%	10.10%
		homemaker	Count	76	35	111
			% within Model	6.60%	7.10%	6.80%
		no benefits not working not seeking	Count	2	0	2
			% within Model	0.20%	0.00%	0.10%
		retired	Count	56	27	83
			% within Model	4.90%	5.50%	5.10%
	Total		Count	1147	490	1637
			% within Model	100.00%	100.00%	100.00%

Recovery				Model		Total
				Allocated	Progressive	
Non Recovered	Employment Status First	Employed	Count	452	171	623
			% within Model	36.40%	35.30%	36.10%
		Unemployed and Seeking Work	Count	447	156	603
			% within Model	36.00%	32.20%	34.90%
		ft students or part students not seeking work	Count	50	25	75
			% within Model	4.00%	5.20%	4.30%
		sick, disabled, incapacity ben, income support, employ & support allowance	Count	129	80	209
			% within Model	10.40%	16.50%	12.10%
		homemaker	Count	104	29	133
			% within Model	8.40%	6.00%	7.70%
		no benefits not working not seeking	Count	1	1	2
			% within Model	0.10%	0.20%	0.10%
		retired	Count	60	23	83
			% within Model	4.80%	4.70%	4.80%
	Total		Count	1243	485	1728
			% within Model	100.00%	100.00%	100.00%
Reliable Deterioration	Employment Status First	Employed	Count	73	20	93
			% within Model	32.60%	27.00%	31.20%
		Unemployed and Seeking Work	Count	88	31	119
			% within Model	39.30%	41.90%	39.90%
		ft students or part students not seeking work	Count	5	2	7
			% within Model	2.20%	2.70%	2.30%
		sick, disabled, incapacity ben, income support, employ & support allowance	Count	27	10	37
			% within Model	12.10%	13.50%	12.40%
		homemaker	Count	20	8	28
			% within Model	8.90%	10.80%	9.40%
		retired	Count	11	3	14
			% within Model	4.90%	4.10%	4.70%
	Total		Count	224	74	298
			% within Model	100.00%	100.00%	100.00%

Recovery				Model		Total
				Allocated	Progressive	
Total	Employment Status First	Employed	Count	2114	1062	3176
			% within Model	48.40%	50.80%	49.10%
		Unemployed and Seeking Work	Count	1196	458	1654
			% within Model	27.40%	21.90%	25.60%
		ft students or part students not seeking work	Count	175	94	269
			% within Model	4.00%	4.50%	4.20%
		sick, disabled, incapacity ben, income support, employ & support allowance	Count	340	207	547
			% within Model	7.80%	9.90%	8.50%
		homemaker	Count	291	136	427
			% within Model	6.70%	6.50%	6.60%
		no benefits not working not seeking	Count	5	1	6
			% within Model	0.10%	0.00%	0.10%
		retired	Count	250	133	383
			% within Model	5.70%	6.40%	5.90%
	Total		Count	4371	2091	6462
			% within Model	100.00%	100.00%	100.00%

Appendix 90: Chi square for first employment status against model and recovery outcome for south cohort

Recovery		Value	df	Asymp. Sig. (2-sided)
Recovered	Pearson Chi-Square	20.903[b]	6	0.002
	Likelihood Ratio	21.845	6	0.001
	Linear-by-Linear Association	1.375	1	0.241
	N of Valid Cases	2799		
Reliable Improvement	Pearson Chi-Square	2.092[c]	6	0.911
	Likelihood Ratio	2.649	6	0.851
	Linear-by-Linear Association	0.735	1	0.391
	N of Valid Cases	1637		
Non Recovered	Pearson Chi-Square	16.447[d]	6	0.012
	Likelihood Ratio	15.878	6	0.014
	Linear-by-Linear Association	0.595	1	0.441
	N of Valid Cases	1728		
Reliable Deterioration	Pearson Chi-Square	1.091[e]	5	0.955
	Likelihood Ratio	1.1	5	0.954
	Linear-by-Linear Association	0.16	1	0.689
	N of Valid Cases	298		
Total	Pearson Chi-Square	28.203[a]	6	0
	Likelihood Ratio	28.518	6	0
	Linear-by-Linear Association	1.178	1	0.278
	N of Valid Cases	6462		

Appendix 91: Cross tabulation of model and outcome for sensitivity analysis

			Embedded Year		Total
			Year 2	Year 4	
Recovery	Recovered	Count	1559	2286	3845
		% within Embedded Year	39.60%	49.20%	44.80%
	Reliable Improvement	Count	1034	1068	2102
		% within Embedded Year	26.30%	23.00%	24.50%
	Non Recovered	Count	1138	1102	2240
		% within Embedded Year	28.90%	23.70%	26.10%
	Reliable Deterioration	Count	201	190	391
		% within Embedded Year	5.10%	4.10%	4.60%
Total		Count	3932	4646	8578
		% within Embedded Year	100.00%	100.00%	100.00%

Appendix 92: Chi square of model and outcome for sensitivity analysis

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	80.021[a]	3	0
Likelihood Ratio	80.228	3	0
Linear-by-Linear Association	68.281	1	0
N of Valid Cases	8578		

Appendix 93: Cross tabulation of model on recovery v non recovery for sensitivity analysis

			Embedded Year		Total
			Year 2	Year 4	
Recovery Non Recovery	Recovered	Count	1559	2286	3845
		% within Embedded Year	39.60%	49.20%	44.80%
	Non Recovered	Count	2373	2360	4733
		% within Embedded Year	60.40%	50.80%	55.20%
Total		Count	3932	4646	8578
		% within Embedded Year	100.00%	100.00%	100.00%

Appendix 94: Chi square of model on recovery v non recovery for sensitivity analysis

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	78.608[a]	1	0		
Continuity Correction[b]	78.223	1	0		
Likelihood Ratio	78.826	1	0		
Fisher's Exact Test				0	0
Linear-by-Linear Association	78.599	1	0		
N of Valid Cases	8578				

Appendix 95: Cross tabulation of model on reliable improvement v non recovery for sensitivity analysis

			Embedded Year		Total
			Year 2	Year 4	
Reliable Improvement Non Recovery	Reliable Improvement	Count	1034	1068	2102
		% within Embedded Year	43.60%	45.30%	44.40%
	Non Recovery	Count	1339	1292	2631
		% within Embedded Year	56.40%	54.70%	55.60%
Total		Count	2373	2360	4733
		% within Embedded Year	100.00%	100.00%	100.00%

Appendix 96: Chi square of model on reliable improvement v non recovery for sensitivity analysis

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1.354[a]	1	0.245		
Continuity Correction[b]	1.287	1	0.257		
Likelihood Ratio	1.354	1	0.245		
Fisher's Exact Test				0.254	0.128
Linear-by-Linear Association	1.354	1	0.245		
N of Valid Cases	4733				

Appendix 97: Cross tabulation of model on reliable deterioration v non recovery for sensitivity analysis

			RD V Non Recovery		Total
			Reliable Deterioration	Non Recovery	
Embedded Year	Year 2	Count	201	2172	2373
		% within RD V Non Recovery	51.40%	50.00%	50.10%
	Year 4	Count	190	2170	2360
		% within RD V Non Recovery	48.60%	50.00%	49.90%
Total		Count	391	4342	4733
		% within RD V Non Recovery	100.00%	100.00%	100.00%

Appendix 98: Chi square of model on reliable deterioration v non recovery for sensitivity analysis

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.275[a]	1	0.6		
Continuity Correction[b]	0.222	1	0.637		
Likelihood Ratio	0.275	1	0.6		
Fisher's Exact Test				0.635	0.319
Linear-by-Linear Association	0.275	1	0.6		
N of Valid Cases	4733				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 194.96.

b. Computed only for a 2x2 table

Appendix 99: Cross tabulation on PHQ Moderate to Severe/Severe on model v outcomes sensitivity analysis

PHQ Severity				Embedded Year		Total
				Year 2	Year 4	
Severe	Recovery	Recovered	Count	303	440	743
			% within Embedded Year	22.70%	30.50%	26.80%
		Reliable Improvement	Count	480	517	997
			% within Embedded Year	36.00%	35.90%	35.90%
		Non Recovered	Count	514	448	962
			% within Embedded Year	38.50%	31.10%	34.70%
		Reliable Deterioration	Count	38	36	74
			% within Embedded Year	2.80%	2.50%	2.70%
	Total		Count	1335	1441	2776
			% within Embedded Year	100.00%	100.00%	100.00%
Moderate to Severe	Recovery	Recovered	Count	459	686	1145
			% within Embedded Year	37.40%	47.70%	43.00%
		Reliable Improvement	Count	359	353	712
			% within Embedded Year	29.30%	24.60%	26.70%
		Non Recovered	Count	334	337	671
			% within Embedded Year	27.20%	23.50%	25.20%
		Reliable Deterioration	Count	75	61	136
			% within Embedded Year	6.10%	4.20%	5.10%
	Total		Count	1227	1437	2664
			% within Embedded Year	100.00%	100.00%	100.00%
Total	Recovery	Recovered	Count	762	1126	1888
			% within Embedded Year	29.70%	39.10%	34.70%
		Reliable Improvement	Count	839	870	1709
			% within Embedded Year	32.70%	30.20%	31.40%
		Non Recovered	Count	848	785	1633
			% within Embedded Year	33.10%	27.30%	30.00%
		Reliable Deterioration	Count	113	97	210
			% within Embedded Year	4.40%	3.40%	3.90%
	Total		Count	2562	2878	5440
			% within Embedded Year	100.00%	100.00%	100.00%

Appendix 100: Chi square on PHQ Moderate to Severe/Severe on model v outcomes for sensitivity analysis

PHQ Severity		Value	df	Asymp. Sig. (2-sided)
Severe	Pearson Chi-Square	27.208[b]	3	0
	Likelihood Ratio	27.317	3	0
	Linear-by-Linear Association	25.045	1	0
	N of Valid Cases	2776		
Moderate to Severe	Pearson Chi-Square	30.142[c]	3	0
	Likelihood Ratio	30.24	3	0
	Linear-by-Linear Association	23.909	1	0
	N of Valid Cases	2664		
Total	Pearson Chi-Square	56.224[a]	3	0
	Likelihood Ratio	56.467	3	0
	Linear-by-Linear Association	50.568	1	0
	N of Valid Cases	5440		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 98.90.
b. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 35.59.
c. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 62.64.

Appendix 101: Cross tabulation on GAD Moderate/Severe on model v outcomes for sensitivity analysis

GAD Severity				Embedded Year		Total
				Year 2	Year 4	
Severe	Recovery	Recovered	Count	668	937	1605
			% within Embedded Year	29.40%	37.70%	33.70%
		Reliable Improvement	Count	795	815	1610
			% within Embedded Year	35.00%	32.80%	33.80%
		Non Recovered	Count	742	683	1425
			% within Embedded Year	32.60%	27.50%	29.90%
		Reliable Deterioration	Count	68	51	119
			% within Embedded Year	3.00%	2.10%	2.50%
	Total		Count	2273	2486	4759
			% within Embedded Year	100.00%	100.00%	100.00%
Moderate	Recovery	Recovered	Count	575	875	1450
			% within Embedded Year	49.60%	59.20%	54.90%
		Reliable Improvement	Count	204	223	427
			% within Embedded Year	17.60%	15.10%	16.20%
		Non Recovered	Count	297	301	598
			% within Embedded Year	25.60%	20.40%	22.70%
		Reliable Deterioration	Count	84	80	164
			% within Embedded Year	7.20%	5.40%	6.20%
	Total		Count	1160	1479	2639
			% within Embedded Year	100.00%	100.00%	100.00%
Total	Recovery	Recovered	Count	1243	1812	3055
			% within Embedded Year	36.20%	45.70%	41.30%
		Reliable Improvement	Count	999	1038	2037
			% within Embedded Year	29.10%	26.20%	27.50%
		Non Recovered	Count	1039	984	2023
			% within Embedded Year	30.30%	24.80%	27.30%
		Reliable Deterioration	Count	152	131	283
			% within Embedded Year	4.40%	3.30%	3.80%
	Total		Count	3433	3965	7398
			% within Embedded Year	100.00%	100.00%	100.00%

Appendix 102: Chi square on GAD Moderate/Severe on model v outcomes for sensitivity analysis

GAD Severity		Value	df	Asymp. Sig. (2-sided)
Severe	Pearson Chi-Square	40.753[b]	3	0
	Likelihood Ratio	40.891	3	0
	Linear-by-Linear Association	37.991	1	0
	N of Valid Cases	4759		
Moderate	Pearson Chi-Square	24.841[c]	3	0
	Likelihood Ratio	24.835	3	0
	Linear-by-Linear Association	22.598	1	0
	N of Valid Cases	2639		
Total	Pearson Chi-Square	71.893[a]	3	0
	Likelihood Ratio	72.111	3	0
	Linear-by-Linear Association	65.053	1	0
	N of Valid Cases	7398		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 131.32.
b. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 56.84.
c. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 72.09.

Appendix 103: Logistical regression on recovered outcome for sensitivity analysis

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Model(1)	0.424	0.056	57.075	1	0	1.527	1.368	1.705
	Gender(1)	-0.024	0.058	0.176	1	0.675	0.976	0.872	1.093
	Disability(1)	-0.025	0.11	0.053	1	0.818	0.975	0.787	1.209
	Age			30.314	5	0			
	Age(1)	0.089	0.113	0.628	1	0.428	1.093	0.877	1.364
	Age(2)	0.227	0.117	3.772	1	0.052	1.255	0.998	1.578
	Age(3)	0.287	0.119	5.835	1	0.016	1.332	1.056	1.681
	Age(4)	0.176	0.127	1.905	1	0.167	1.192	0.929	1.53
	Age(5)	1.01	0.208	23.504	1	0	2.745	1.825	4.129
	Employment			160.801	5	0			
	Employment(1)	-0.725	0.073	99.336	1	0	0.485	0.42	0.559
	Employment(2)	-0.259	0.129	4.06	1	0.044	0.771	0.599	0.993
	Employment(3)	-0.895	0.091	96.315	1	0	0.409	0.342	0.489
	Employment(4)	-0.488	0.119	16.762	1	0	0.614	0.486	0.775
	Employment(5)	-0.73	0.181	16.297	1	0	0.482	0.338	0.687
	PHQ			151.663	4	0			
	PHQ(1)	-0.159	0.224	0.502	1	0.479	0.853	0.55	1.324
	PHQ(2)	-0.661	0.211	9.815	1	0.002	0.516	0.342	0.781
	PHQ(3)	-1.034	0.21	24.188	1	0	0.356	0.235	0.537
	PHQ(4)	-1.339	0.213	39.578	1	0	0.262	0.173	0.398
	GAD			124.811	3	0			
	GAD(1)	0.057	0.244	0.055	1	0.815	1.059	0.656	1.709
	GAD(2)	-0.455	0.236	3.738	1	0.053	0.634	0.4	1.006
	GAD(3)	-0.914	0.235	15.166	1	0	0.401	0.253	0.635
	Discharge			1548.424	5	0			
	Discharge(1)	-2.774	0.085	1061.675	1	0	0.062	0.053	0.074
	Discharge(2)	-2.434	0.206	138.909	1	0	0.088	0.059	0.131
	Discharge(3)	-2.122	0.123	300.067	1	0	0.12	0.094	0.152
	Discharge(4)	-3.434	0.217	251.076	1	0	0.032	0.021	0.049
	Discharge(5)	-1.329	0.14	89.779	1	0	0.265	0.201	0.349
	Constant	2.181	0.328	44.166	1	0	8.853		

a. Variable(s) entered on step 1: Model, Gender, Disability, Age, Employment, PHQ, GAD, Discharge.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 104: Logistical regression on recovered outcome for allocated model

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1a	Gender(1)	-0.07	0.084	0.702	1	0.402	0.932	0.791	1.099
	Disability(1)	0.095	0.202	0.221	1	0.638	1.1	0.74	1.634
	Age			28.038	5	0			
	Age(1)	-0.027	0.185	0.021	1	0.886	0.974	0.678	1.4
	Age(2)	0.241	0.192	1.572	1	0.21	1.273	0.873	1.856
	Age(3)	0.283	0.194	2.126	1	0.145	1.327	0.907	1.94
	Age(4)	0.248	0.203	1.492	1	0.222	1.281	0.861	1.907
	Age(5)	1.271	0.3	17.939	1	0	3.564	1.979	6.416
	Employment			83.242	5	0			
	Employment(1)	-0.596	0.101	35.079	1	0	0.551	0.452	0.671
	Employment(2)	-0.055	0.191	0.083	1	0.773	0.947	0.652	1.375
	Employment(3)	-1.135	0.154	54.133	1	0	0.321	0.238	0.435
	Employment(4)	-0.606	0.171	12.65	1	0	0.545	0.39	0.762
	Employment(5)	-0.975	0.264	13.652	1	0	0.377	0.225	0.633
	PHQ			94.405	4	0			
	PHQ(1)	-0.379	0.331	1.31	1	0.252	0.684	0.357	1.31
	PHQ(2)	-0.779	0.313	6.174	1	0.013	0.459	0.248	0.848
	PHQ(3)	-1.36	0.313	18.863	1	0	0.257	0.139	0.474
	PHQ(4)	-1.582	0.316	25.037	1	0	0.206	0.111	0.382
	GAD			52.911	3	0			
	GAD(1)	0.328	0.303	1.172	1	0.279	1.388	0.767	2.514
	GAD(2)	-0.152	0.288	0.277	1	0.599	0.859	0.488	1.512
	GAD(3)	-0.593	0.287	4.263	1	0.039	0.553	0.315	0.97
	Discharge			592.094	5	0			
	Discharge(1)	-2.401	0.12	403.088	1	0	0.091	0.072	0.115
	Discharge(2)	-2.578	0.327	62.111	1	0	0.076	0.04	0.144
	Discharge(3)	-1.637	0.174	88.486	1	0	0.195	0.138	0.274
	Discharge(4)	-3.08	0.296	108.146	1	0	0.046	0.026	0.082
	Discharge(5)	-1.475	0.227	42.169	1	0	0.229	0.147	0.357
	Constant	2.044	0.451	20.538	1	0	7.719		

a. Variable(s) entered on step 1: Gender, Disability, Age, Employment, PHQ, GAD, Discharge.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 105: Logistical regression on reliably improved outcome for allocated model

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1a	Gender(1)	0.037	0.081	0.212	1	0.645	1.038	0.886	1.216
	Disability(1)	-0.064	0.193	0.11	1	0.741	0.938	0.643	1.369
	Age			10.827	5	0.055			
	Age(1)	0.131	0.179	0.535	1	0.464	1.14	0.803	1.618
	Age(2)	0.001	0.186	0	1	0.997	1.001	0.696	1.44
	Age(3)	-0.056	0.188	0.09	1	0.764	0.945	0.654	1.366
	Age(4)	-0.018	0.197	0.008	1	0.928	0.982	0.668	1.446
	Age(5)	-0.739	0.317	5.442	1	0.02	0.477	0.256	0.889
	Employment			4.543	5	0.474			
	Employment(1)	-0.094	0.096	0.961	1	0.327	0.91	0.755	1.098
	Employment(2)	-0.089	0.194	0.211	1	0.646	0.915	0.626	1.337
	Employment(3)	0.111	0.131	0.727	1	0.394	1.118	0.865	1.444
	Employment(4)	-0.029	0.163	0.032	1	0.857	0.971	0.705	1.337
	Employment(5)	0.376	0.271	1.928	1	0.165	1.456	0.857	2.476
	PHQ			79.629	4	0			
	PHQ(1)	0.587	0.5	1.381	1	0.24	1.799	0.676	4.792
	PHQ(2)	0.885	0.477	3.448	1	0.063	2.424	0.952	6.172
	PHQ(3)	1.628	0.473	11.865	1	0.001	5.095	2.017	12.867
	PHQ(4)	1.69	0.474	12.72	1	0	5.418	2.141	13.714
	GAD			91.985	3	0			
	GAD(1)	1.161	0.74	2.457	1	0.117	3.192	0.748	13.626
	GAD(2)	1.837	0.725	6.416	1	0.011	6.275	1.515	25.985
	GAD(3)	2.494	0.723	11.896	1	0.001	12.11	2.935	49.964
	Discharge			14.942	5	0.011			
	Discharge(1)	0.262	0.094	7.727	1	0.005	1.3	1.08	1.564
	Discharge(2)	-0.305	0.216	1.992	1	0.158	0.737	0.482	1.126
	Discharge(3)	0.093	0.161	0.333	1	0.564	1.098	0.8	1.506
	Discharge(4)	-0.203	0.159	1.619	1	0.203	0.816	0.597	1.116
	Discharge(5)	-0.101	0.215	0.22	1	0.639	0.904	0.593	1.378
	Constant	-4.681	0.877	28.509	1	0	0.009		

a. Variable(s) entered on step 1: Gender, Disability, Age, Employment, PHQ, GAD, Discharge.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 106: Logistical regression on no change outcome for allocated model

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1a	Gender(1)	0.086	0.083	1.078	1	0.299	1.09	0.926	1.282
	Disability(1)	0.032	0.194	0.027	1	0.869	1.033	0.706	1.511
	Age			8.171	5	0.147			
	Age(1)	-0.132	0.171	0.589	1	0.443	0.877	0.627	1.227
	Age(2)	-0.308	0.179	2.948	1	0.086	0.735	0.517	1.045
	Age(3)	-0.234	0.181	1.677	1	0.195	0.791	0.555	1.128
	Age(4)	-0.209	0.192	1.185	1	0.276	0.811	0.557	1.182
	Age(5)	-0.767	0.317	5.866	1	0.015	0.464	0.25	0.864
	Employment			43.155	5	0			
	Employment(1)	0.494	0.097	25.816	1	0	1.638	1.354	1.982
	Employment(2)	0.206	0.191	1.164	1	0.281	1.229	0.845	1.786
	Employment(3)	0.63	0.133	22.313	1	0	1.877	1.445	2.437
	Employment(4)	0.615	0.161	14.626	1	0	1.85	1.35	2.535
	Employment(5)	0.688	0.272	6.411	1	0.011	1.99	1.168	3.391
	PHQ			15.966	4	0.003			
	PHQ(1)	0.075	0.389	0.037	1	0.848	1.077	0.502	2.311
	PHQ(2)	0.431	0.366	1.388	1	0.239	1.54	0.751	3.156
	PHQ(3)	0.425	0.366	1.354	1	0.245	1.53	0.747	3.133
	PHQ(4)	0.686	0.367	3.5	1	0.061	1.986	0.968	4.077
	GAD			8.026	3	0.045			
	GAD(1)	-0.568	0.337	2.832	1	0.092	0.567	0.292	1.098
	GAD(2)	-0.166	0.32	0.269	1	0.604	0.847	0.452	1.586
	GAD(3)	-0.188	0.319	0.348	1	0.556	0.829	0.444	1.548
	Discharge			402.548	5	0			
	Discharge(1)	1.666	0.095	307.442	1	0	5.293	4.394	6.377
	Discharge(2)	1.622	0.188	74.153	1	0	5.065	3.501	7.327
	Discharge(3)	1.373	0.149	84.504	1	0	3.946	2.945	5.287
	Discharge(4)	1.832	0.146	157.951	1	0	6.249	4.696	8.317
	Discharge(5)	1.523	0.194	61.685	1	0	4.586	3.136	6.707
	Constant	-2.167	0.503	18.53	1	0	0.115		

a. Variable(s) entered on step 1: Gender, Disability, Age, Employment, PHQ, GAD, Discharge.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 107: Logistical regression on reliably deteriorated outcome for allocated model

								Lower	Upper
Step 1a	Gender(1)	-0.211	0.165	1.642	1	0.2	0.81	0.586	1.118
	Disability(1)	0.182	0.369	0.243	1	0.622	1.2	0.582	2.475
	Age			2.37	5	0.796			
	Age(1)	0.229	0.333	0.471	1	0.493	1.257	0.654	2.415
	Age(2)	0.445	0.344	1.671	1	0.196	1.56	0.795	3.063
	Age(3)	0.259	0.349	0.551	1	0.458	1.296	0.654	2.571
	Age(4)	0.166	0.379	0.192	1	0.661	1.181	0.562	2.481
	Age(5)	0.295	0.61	0.234	1	0.629	1.343	0.406	4.439
	Employment			24.26	5	0			
	Employment(1)	0.772	0.192	16.176	1	0	2.165	1.486	3.154
	Employment(2)	0.128	0.379	0.114	1	0.735	1.137	0.541	2.39
	Employment(3)	0.944	0.251	14.132	1	0	2.571	1.571	4.206
	Employment(4)	0.156	0.342	0.208	1	0.648	1.169	0.598	2.286
	Employment(5)	-0.169	0.586	0.083	1	0.773	0.844	0.268	2.662
	PHQ			25.735	4	0			
	PHQ(1)	0.16	0.585	0.075	1	0.785	1.173	0.373	3.692
	PHQ(2)	0.044	0.559	0.006	1	0.937	1.045	0.35	3.124
	PHQ(3)	-0.21	0.564	0.138	1	0.71	0.811	0.269	2.448
	PHQ(4)	-1.162	0.589	3.891	1	0.049	0.313	0.099	0.993
	GAD			46.822	3	0			
	GAD(1)	-0.426	0.45	0.898	1	0.343	0.653	0.27	1.577
	GAD(2)	-0.952	0.44	4.676	1	0.031	0.386	0.163	0.915
	GAD(3)	-1.868	0.451	17.117	1	0	0.154	0.064	0.374
	Discharge			130.793	5	0			
	Discharge(1)	1.58	0.206	58.944	1	0	4.855	3.244	7.267
	Discharge(2)	2.771	0.3	85.356	1	0	15.973	8.874	28.753
	Discharge(3)	1.708	0.291	34.325	1	0	5.516	3.115	9.766
	Discharge(4)	2.379	0.26	83.41	1	0	10.794	6.478	17.984
	Discharge(5)	1.34	0.411	10.635	1	0.001	3.82	1.707	8.549
	Constant	-3.003	0.766	15.376	1	0	0.05		

a. Variable(s) entered on step 1: Gender, Disability, Age, Employment, PHQ, GAD, Discharge.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 108: Logistical regression on recovered outcome for progressive model

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1a	Gender(1)	0.01	0.08	0.015	1	0.902	1.01	0.863	1.182
	Disability(1)	-0.072	0.134	0.287	1	0.592	0.931	0.715	1.211
	Age			8.317	5	0.14			
	Age(1)	0.124	0.147	0.714	1	0.398	1.132	0.849	1.511
	Age(2)	0.152	0.151	1.004	1	0.316	1.164	0.865	1.566
	Age(3)	0.229	0.155	2.199	1	0.138	1.258	0.929	1.703
	Age(4)	0.054	0.17	0.1	1	0.752	1.055	0.757	1.471
	Age(5)	0.715	0.305	5.504	1	0.019	2.045	1.125	3.718
	Employment			91.161	5	0			
	Employment(1)	-0.878	0.106	69.13	1	0	0.416	0.338	0.511
	Employment(2)	-0.425	0.177	5.761	1	0.016	0.654	0.462	0.925
	Employment(3)	-0.802	0.117	47.144	1	0	0.449	0.357	0.564
	Employment(4)	-0.386	0.17	5.151	1	0.023	0.68	0.487	0.949
	Employment(5)	-0.533	0.256	4.342	1	0.037	0.587	0.356	0.969
	PHQ			64.99	4	0			
	PHQ(1)	0.044	0.304	0.021	1	0.886	1.044	0.576	1.894
	PHQ(2)	-0.571	0.284	4.051	1	0.044	0.565	0.324	0.985
	PHQ(3)	-0.745	0.283	6.941	1	0.008	0.475	0.273	0.826
	PHQ(4)	-1.123	0.287	15.362	1	0	0.325	0.185	0.57
	GAD			75.24	3	0			
	GAD(1)	-0.393	0.419	0.876	1	0.349	0.675	0.297	1.536
	GAD(2)	-0.947	0.409	5.357	1	0.021	0.388	0.174	0.865
	GAD(3)	-1.433	0.408	12.331	1	0	0.239	0.107	0.531
	Discharge			949.005	5	0			
	Discharge(1)	-3.13	0.123	650.031	1	0	0.044	0.034	0.056
	Discharge(2)	-2.35	0.271	74.929	1	0	0.095	0.056	0.162
	Discharge(3)	-2.535	0.173	215.567	1	0	0.079	0.056	0.111
	Discharge(4)	-3.77	0.319	139.896	1	0	0.023	0.012	0.043
	Discharge(5)	-1.245	0.182	46.925	1	0	0.288	0.202	0.411
	Constant	3.017	0.509	35.133	1	0	20.423		

a. Variable(s) entered on step 1: Gender, Disability, Age, Employment, PHQ, GAD, Discharge.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 109: Logistical regression on reliably improved outcome for progressive model

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1a	Gender(1)	0	0.079	0	1	0.997	1	0.856	1.168
	Disability(1)	-0.042	0.128	0.111	1	0.739	0.958	0.746	1.231
	Age			21.099	5	0.001			
	Age(1)	0.246	0.14	3.104	1	0.078	1.279	0.973	1.682
	Age(2)	-0.122	0.146	0.699	1	0.403	0.885	0.665	1.178
	Age(3)	-0.106	0.148	0.511	1	0.475	0.899	0.672	1.203
	Age(4)	0.195	0.162	1.452	1	0.228	1.215	0.885	1.668
	Age(5)	-0.423	0.323	1.713	1	0.191	0.655	0.348	1.234
	Employment			6.299	5	0.278			
	Employment(1)	0.213	0.1	4.561	1	0.033	1.238	1.018	1.505
	Employment(2)	0.255	0.176	2.107	1	0.147	1.291	0.914	1.822
	Employment(3)	0.049	0.112	0.194	1	0.66	1.051	0.843	1.31
	Employment(4)	0.036	0.17	0.046	1	0.83	1.037	0.743	1.447
	Employment(5)	0.257	0.268	0.925	1	0.336	1.293	0.766	2.185
	PHQ			66.157	4	0			
	PHQ(1)	0.73	0.548	1.777	1	0.182	2.075	0.71	6.068
	PHQ(2)	1.326	0.525	6.389	1	0.011	3.767	1.347	10.535
	PHQ(3)	1.684	0.521	10.433	1	0.001	5.386	1.939	14.961
	PHQ(4)	1.966	0.522	14.183	1	0	7.145	2.568	19.883
	GAD			145.298	3	0			
	GAD(1)	-0.153	0.765	0.04	1	0.841	0.858	0.192	3.841
	GAD(2)	1.407	0.728	3.731	1	0.053	4.083	0.979	17.019
	GAD(3)	2.201	0.726	9.183	1	0.002	9.037	2.176	37.528
	Discharge			44.751	5	0			
	Discharge(1)	0.485	0.091	28.315	1	0	1.624	1.358	1.941
	Discharge(2)	0.139	0.237	0.346	1	0.557	1.149	0.723	1.827
	Discharge(3)	0.623	0.146	18.208	1	0	1.865	1.401	2.483
	Discharge(4)	0.008	0.162	0.003	1	0.959	1.008	0.734	1.385
	Discharge(5)	-0.241	0.218	1.228	1	0.268	0.786	0.513	1.204
	Constant	-4.939	0.901	30.083	1	0	0.007		

a. Variable(s) entered on step 1: Gender, Disability, Age, Employment, PHQ, GAD, Discharge.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 110: Logistical regression on no change outcome for progressive model

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1a	Gender(1)	0.047	0.084	0.307	1	0.579	1.048	0.888	1.237
	Disability(1)	0.076	0.134	0.325	1	0.569	1.079	0.83	1.404
	Age			11.055	5	0.05			
	Age(1)	-0.325	0.143	5.153	1	0.023	0.722	0.546	0.957
	Age(2)	0.003	0.146	0	1	0.984	1.003	0.754	1.334
	Age(3)	-0.114	0.151	0.571	1	0.45	0.892	0.664	1.199
	Age(4)	-0.234	0.171	1.877	1	0.171	0.791	0.566	1.106
	Age(5)	-0.428	0.348	1.511	1	0.219	0.652	0.329	1.29
	Employment			35.796	5	0			
	Employment(1)	0.473	0.106	20.031	1	0	1.605	1.305	1.975
	Employment(2)	0.322	0.182	3.118	1	0.077	1.38	0.965	1.972
	Employment(3)	0.631	0.119	28.275	1	0	1.879	1.489	2.371
	Employment(4)	0.365	0.174	4.378	1	0.036	1.44	1.023	2.027
	Employment(5)	0.312	0.301	1.075	1	0.3	1.367	0.757	2.468
	PHQ			5.844	4	0.211			
	PHQ(1)	0.268	0.363	0.544	1	0.461	1.307	0.641	2.664
	PHQ(2)	0.518	0.342	2.292	1	0.13	1.679	0.858	3.285
	PHQ(3)	0.515	0.341	2.283	1	0.131	1.674	0.858	3.268
	PHQ(4)	0.599	0.344	3.022	1	0.082	1.82	0.927	3.574
	GAD			4.534	3	0.209			
	GAD(1)	1.145	0.564	4.114	1	0.043	3.142	1.039	9.498
	GAD(2)	1.183	0.557	4.515	1	0.034	3.264	1.096	9.717
	GAD(3)	1.15	0.555	4.29	1	0.038	3.159	1.064	9.381
	Discharge			694.622	5	0			
	Discharge(1)	2.287	0.097	557.939	1	0	9.841	8.141	11.898
	Discharge(2)	2.191	0.208	111.305	1	0	8.947	5.955	13.442
	Discharge(3)	2.029	0.141	207.527	1	0	7.61	5.774	10.03
	Discharge(4)	2.491	0.15	275.345	1	0	12.077	8.998	16.208
	Discharge(5)	1.839	0.182	101.913	1	0	6.293	4.403	8.994
	Constant	-4.166	0.66	39.88	1	0	0.016		

a. Variable(s) entered on step 1: Gender, Disability, Age, Employment, PHQ, GAD, Discharge.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 111: Logistical regression on reliably deteriorated outcome for progressive model

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1a	Gender(1)	-0.327	0.178	3.375	1	0.066	0.721	0.509	1.022
	Disability(1)	0.144	0.261	0.302	1	0.583	1.154	0.692	1.926
	Age			1.187	5	0.946			
	Age(1)	-0.182	0.28	0.423	1	0.515	0.833	0.481	1.443
	Age(2)	-0.046	0.29	0.025	1	0.874	0.955	0.541	1.686
	Age(3)	0.019	0.299	0.004	1	0.949	1.019	0.567	1.831
	Age(4)	0.097	0.331	0.087	1	0.768	1.102	0.577	2.107
	Age(5)	0.189	0.704	0.072	1	0.788	1.208	0.304	4.803
	Employment			27.958	5	0			
	Employment(1)	0.975	0.214	20.707	1	0	2.651	1.742	4.034
	Employment(2)	-0.078	0.4	0.038	1	0.846	0.925	0.423	2.025
	Employment(3)	0.937	0.248	14.212	1	0	2.551	1.568	4.152
	Employment(4)	0.607	0.344	3.122	1	0.077	1.836	0.936	3.6
	Employment(5)	0.096	0.66	0.021	1	0.885	1.1	0.302	4.01
	PHQ			24.92	4	0			
	PHQ(1)	-1.113	0.451	6.092	1	0.014	0.329	0.136	0.795
	PHQ(2)	-0.889	0.4	4.95	1	0.026	0.411	0.188	0.9
	PHQ(3)	-1.347	0.409	10.824	1	0.001	0.26	0.117	0.58
	PHQ(4)	-1.957	0.446	19.276	1	0	0.141	0.059	0.339
	GAD			75.034	3	0			
	GAD(1)	-0.92	0.5	3.379	1	0.066	0.399	0.15	1.063
	GAD(2)	-1.171	0.487	5.782	1	0.016	0.31	0.119	0.805
	GAD(3)	-2.777	0.508	29.899	1	0	0.062	0.023	0.168
	Discharge			158.628	5	0			
	Discharge(1)	2.594	0.229	127.976	1	0	13.382	8.538	20.975
	Discharge(2)	2.359	0.423	31.064	1	0	10.586	4.617	24.27
	Discharge(3)	2.161	0.308	49.314	1	0	8.683	4.75	15.874
	Discharge(4)	3.215	0.297	117.237	1	0	24.908	13.918	44.576
	Discharge(5)	1.531	0.475	10.395	1	0.001	4.621	1.822	11.717
	Constant	-2.079	0.663	9.839	1	0.002	0.125		

a. Variable(s) entered on step 1: Gender, Disability, Age, Employment, PHQ, GAD, Discharge.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 112: Logistical regression on recovered outcome on PHQ moderate for sensitivity analysis

Step 1[a]	MODEL(1)	0.244	0.111	4.85	1	0.028	1.277	1.027	1.587
	GENDER(1)	-0.06	0.114	0.274	1	0.601	0.942	0.753	1.179
	DISABILITY(1)	0.177	0.251	0.5	1	0.48	1.194	0.73	1.953
	AGEREGRESSION			8.077	5	0.152			
	AGEREGRESSION(1)	0.283	0.204	1.928	1	0.165	1.327	0.89	1.98
	AGEREGRESSION(2)	0.327	0.215	2.321	1	0.128	1.387	0.91	2.114
	AGEREGRESSION(3)	0.384	0.225	2.912	1	0.088	1.469	0.945	2.283
	AGEREGRESSION(4)	0.567	0.243	5.455	1	0.02	1.764	1.096	2.839
	AGEREGRESSION(5)	1.011	0.41	6.087	1	0.014	2.748	1.231	6.133
	EMPLOYMENT			38.569	5	0			
	EMPLOYMENT(1)	-0.509	0.154	10.851	1	0.001	0.601	0.444	0.814
	EMPLOYMENT(2)	-0.094	0.229	0.166	1	0.683	0.911	0.581	1.428
	EMPLOYMENT(3)	-1.09	0.205	28.276	1	0	0.336	0.225	0.503
	EMPLOYMENT(4)	-0.616	0.246	6.292	1	0.012	0.54	0.334	0.874
	EMPLOYMENT(5)	-0.809	0.354	5.234	1	0.022	0.445	0.223	0.891
	GAD			49.18	3	0			
	GAD(1)	-0.208	0.322	0.418	1	0.518	0.812	0.433	1.525
	GAD(2)	-0.623	0.311	4.012	1	0.045	0.536	0.292	0.987
	GAD(3)	-1.181	0.314	14.104	1	0	0.307	0.166	0.569
	DISCHARGE			407.887	5	0			
	DISCHARGE(1)	-2.732	0.158	297.91	1	0	0.065	0.048	0.089
	DISCHARGE(2)	-1.754	0.388	20.4	1	0	0.173	0.081	0.37
	DISCHARGE(3)	-1.91	0.217	77.208	1	0	0.148	0.097	0.227
	DISCHARGE(4)	-3.216	0.391	67.516	1	0	0.04	0.019	0.086
	DISCHARGE(5)	-1.781	0.299	35.387	1	0	0.169	0.094	0.303
	Constant	1.644	0.363	20.52	1	0	5.174		

a. Variable(s) entered on step 1: MODEL, GENDER, DISABILITY, AGEREGRESSION, EMPLOYMENT, GAD, DISCHARGE.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 113: Logistical regression on reliably improved outcome on PHQ moderate for sensitivity analysis

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1a	MODEL(1)	-0.016	0.141	0.012	1	0.912	0.985	0.747	1.298
	GENDER(1)	-0.234	0.149	2.47	1	0.116	0.791	0.591	1.06
	DISABILITY(1)	0.201	0.332	0.365	1	0.546	1.222	0.637	2.345
	AGEREGRESSION			4.669	5	0.458			
	AGEREGRESSION(1)	0.066	0.256	0.067	1	0.796	1.068	0.647	1.764
	AGEREGRESSION(2)	-0.276	0.28	0.969	1	0.325	0.759	0.438	1.314
	AGEREGRESSION(3)	0.07	0.287	0.059	1	0.808	1.072	0.61	1.883
	AGEREGRESSION(4)	-0.263	0.324	0.657	1	0.418	0.769	0.407	1.452
	AGEREGRESSION(5)	-0.257	0.556	0.213	1	0.644	0.774	0.26	2.299
	EMPLOYMENT			8.005	5	0.156			
	EMPLOYMENT(1)	-0.01	0.197	0.002	1	0.961	0.99	0.673	1.458
	EMPLOYMENT(2)	0.043	0.293	0.021	1	0.884	1.044	0.588	1.852
	EMPLOYMENT(3)	0.471	0.251	3.521	1	0.061	1.602	0.979	2.622
	EMPLOYMENT(4)	-0.662	0.363	3.321	1	0.068	0.516	0.253	1.051
	EMPLOYMENT(5)	0.38	0.495	0.59	1	0.443	1.463	0.554	3.86
	GAD			123.765	3	0			
	GAD(1)	17.373	4387.093	0	1	0.997	35057651	0	.
	GAD(2)	18.954	4387.093	0	1	0.997	1.7E+08	0	.
	GAD(3)	20.178	4387.093	0	1	0.996	5.8E+08	0	.
	DISCHARGE			35.886	5	0			
	DISCHARGE(1)	0.882	0.163	29.248	1	0	2.416	1.755	3.327
	DISCHARGE(2)	-0.414	0.757	0.299	1	0.585	0.661	0.15	2.914
	DISCHARGE(3)	0.739	0.279	7.03	1	0.008	2.094	1.213	3.617
	DISCHARGE(4)	0.845	0.333	6.458	1	0.011	2.328	1.213	4.467
	DISCHARGE(5)	0.032	0.464	0.005	1	0.945	1.033	0.416	2.565
	Constant	-21.295	4387.093	0	1	0.996	0		

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 114: Logistical regression on no change outcome on PHQ moderate for sensitivity analysis

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1a	MODEL(1)	-0.249	0.121	4.218	1	0.04	0.78	0.615	0.989
	GENDER(1)	0.199	0.125	2.537	1	0.111	1.221	0.955	1.56
	DISABILITY(1)	-0.18	0.3	0.358	1	0.55	0.836	0.464	1.506
	AGEREGRESSION			12.069	5	0.034			
	AGEREGRESSION(1)	-0.203	0.204	0.99	1	0.32	0.816	0.547	1.218
	AGEREGRESSION(2)	-0.1	0.217	0.214	1	0.643	0.904	0.591	1.384
	AGEREGRESSION(3)	-0.54	0.237	5.21	1	0.022	0.583	0.367	0.927
	AGEREGRESSION(4)	-0.408	0.257	2.51	1	0.113	0.665	0.402	1.101
	AGEREGRESSION(5)	-1.306	0.5	6.83	1	0.009	0.271	0.102	0.721
	EMPLOYMENT			21.572	5	0.001			
	EMPLOYMENT(1)	0.134	0.163	0.677	1	0.411	1.143	0.831	1.574
	EMPLOYMENT(2)	0.152	0.237	0.409	1	0.523	1.164	0.731	1.853
	EMPLOYMENT(3)	0.54	0.221	5.993	1	0.014	1.717	1.114	2.646
	EMPLOYMENT(4)	0.907	0.242	14.063	1	0	2.476	1.542	3.978
	EMPLOYMENT(5)	0.987	0.414	5.69	1	0.017	2.683	1.192	6.035
	GAD			18.234	3	0			
	GAD(1)	0.483	0.39	1.53	1	0.216	1.62	0.754	3.481
	GAD(2)	0.754	0.38	3.929	1	0.047	2.126	1.008	4.48
	GAD(3)	0.182	0.388	0.22	1	0.639	1.2	0.561	2.565
	DISCHARGE			203.82	5	0			
	DISCHARGE(1)	1.836	0.141	169.845	1	0	6.274	4.76	8.27
	DISCHARGE(2)	0.607	0.447	1.842	1	0.175	1.835	0.764	4.411
	DISCHARGE(3)	1.488	0.221	45.386	1	0	4.428	2.872	6.826
	DISCHARGE(4)	1.673	0.27	38.508	1	0	5.329	3.141	9.039
	DISCHARGE(5)	1.78	0.294	36.668	1	0	5.933	3.334	10.557
	Constant	-2.396	0.426	31.667	1	0	0.091		

a. Variable(s) entered on step 1: MODEL, GENDER, DISABILITY, AGEREGRESSION, EMPLOYMENT, GAD, DISCHARGE.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 115: Logistical regression on reliably deteriorated outcome on PHQ moderate for sensitivity analysis

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1a	MODEL(1)	-0.069	0.2	0.119	1	0.731	0.933	0.63	1.382
	GENDER(1)	0.075	0.204	0.134	1	0.714	1.078	0.722	1.609
	DISABILITY(1)	-0.424	0.491	0.746	1	0.388	0.654	0.25	1.713
	AGEREGRESSION			3.349	5	0.646			
	AGEREGRESSION(1)	-0.244	0.322	0.571	1	0.45	0.784	0.417	1.474
	AGEREGRESSION(2)	0.039	0.339	0.013	1	0.908	1.04	0.535	2.021
	AGEREGRESSION(3)	0.151	0.35	0.187	1	0.666	1.163	0.586	2.309
	AGEREGRESSION(4)	-0.102	0.397	0.067	1	0.797	0.903	0.414	1.966
	AGEREGRESSION(5)	0.742	0.725	1.047	1	0.306	2.099	0.507	8.693
	EMPLOYMENT			26.654	5	0			
	EMPLOYMENT(1)	0.992	0.234	18.011	1	0	2.697	1.706	4.266
	EMPLOYMENT(2)	0.048	0.44	0.012	1	0.914	1.049	0.443	2.484
	EMPLOYMENT(3)	1.035	0.321	10.385	1	0.001	2.814	1.5	5.279
	EMPLOYMENT(4)	0.374	0.446	0.705	1	0.401	1.454	0.607	3.481
	EMPLOYMENT(5)	-0.878	0.832	1.112	1	0.292	0.416	0.081	2.125
	GAD			25.35	3	0			
	GAD(1)	-0.689	0.425	2.623	1	0.105	0.502	0.218	1.156
	GAD(2)	-1.597	0.431	13.725	1	0	0.202	0.087	0.471
	GAD(3)	-1.527	0.437	12.233	1	0	0.217	0.092	0.511
	DISCHARGE			117.161	5	0			
	DISCHARGE(1)	2.261	0.257	77.524	1	0	9.593	5.799	15.868
	DISCHARGE(2)	3.294	0.445	54.784	1	0	26.961	11.269	64.506
	DISCHARGE(3)	1.932	0.351	30.334	1	0	6.903	3.471	13.728
	DISCHARGE(4)	2.886	0.352	67.109	1	0	17.923	8.985	35.751
	DISCHARGE(5)	1.797	0.489	13.475	1	0	6.029	2.31	15.735
	Constant	-3.019	0.516	34.232	1	0	0.049		

a. Variable(s) entered on step 1: MODEL, GENDER, DISABILITY, AGEREGRESSION, EMPLOYMENT, GAD, DISCHARGE.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 116: Logistical regression on recovered outcome on PHQ moderate/severe for sensitivity analysis

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1a	MODEL(1)	0.584	0.098	35.493	1	0	1.794	1.48	2.174
	GENDER(1)	-0.027	0.101	0.071	1	0.79	0.974	0.799	1.186
	DISABILITY(1)	-0.104	0.194	0.287	1	0.592	0.901	0.616	1.318
	AGEREGRESSION			8.744	5	0.12			
	AGEREGRESSION(1)	0.344	0.185	3.449	1	0.063	1.411	0.981	2.028
	AGEREGRESSION(2)	0.26	0.192	1.83	1	0.176	1.297	0.89	1.892
	AGEREGRESSION(3)	0.242	0.197	1.514	1	0.219	1.274	0.866	1.872
	AGEREGRESSION(4)	0.145	0.212	0.466	1	0.495	1.156	0.762	1.753
	AGEREGRESSION(5)	0.936	0.38	6.063	1	0.014	2.549	1.21	5.367
	EMPLOYMENT			33.456	5	0			
	EMPLOYMENT(1)	-0.677	0.125	29.378	1	0	0.508	0.398	0.649
	EMPLOYMENT(2)	-0.228	0.214	1.14	1	0.286	0.796	0.523	1.21
	EMPLOYMENT(3)	-0.513	0.153	11.192	1	0.001	0.599	0.443	0.809
	EMPLOYMENT(4)	-0.201	0.209	0.925	1	0.336	0.818	0.542	1.232
	EMPLOYMENT(5)	-0.406	0.368	1.222	1	0.269	0.666	0.324	1.369
	GAD			19.972	3	0			
	GAD(1)	0.427	0.468	0.835	1	0.361	1.533	0.613	3.832
	GAD(2)	-0.013	0.445	0.001	1	0.977	0.987	0.413	2.363
	GAD(3)	-0.29	0.442	0.431	1	0.511	0.748	0.315	1.779
	DISCHARGE			493.006	5	0			
	DISCHARGE(1)	-2.894	0.158	333.671	1	0	0.055	0.041	0.076
	DISCHARGE(2)	-3.055	0.475	41.389	1	0	0.047	0.019	0.12
	DISCHARGE(3)	-2.398	0.247	94.013	1	0	0.091	0.056	0.148
	DISCHARGE(4)	-3.282	0.373	77.359	1	0	0.038	0.018	0.078
	DISCHARGE(5)	-1.086	0.226	23.159	1	0	0.338	0.217	0.525
	Constant	0.401	0.476	0.708	1	0.4	1.493		

a. Variable(s) entered on step 1: MODEL, GENDER, DISABILITY, AGEREGRESSION, EMPLOYMENT, GAD, DISCHARGE.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 117: Logistical regression on reliably improved outcome on PHQ moderate/severe for sensitivity analysis

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1a	MODEL(1)	-0.346	0.093	13.743	1	0	0.708	0.589	0.85
	GENDER(1)	0.037	0.097	0.15	1	0.699	1.038	0.859	1.255
	DISABILITY(1)	0.192	0.184	1.095	1	0.295	1.212	0.845	1.737
	AGEREGRESSION			5.378	5	0.371			
	AGEREGRESSION(1)	0.057	0.171	0.111	1	0.739	1.059	0.758	1.479
	AGEREGRESSION(2)	-0.048	0.179	0.071	1	0.789	0.953	0.672	1.353
	AGEREGRESSION(3)	-0.142	0.186	0.581	1	0.446	0.868	0.602	1.25
	AGEREGRESSION(4)	0.039	0.203	0.036	1	0.849	1.039	0.698	1.547
	AGEREGRESSION(5)	-0.625	0.397	2.482	1	0.115	0.535	0.246	1.165
	EMPLOYMENT			2.593	5	0.762			
	EMPLOYMENT(1)	0.014	0.116	0.014	1	0.905	1.014	0.807	1.274
	EMPLOYMENT(2)	-0.004	0.201	0	1	0.984	0.996	0.671	1.478
	EMPLOYMENT(3)	-0.206	0.15	1.874	1	0.171	0.814	0.606	1.093
	EMPLOYMENT(4)	-0.138	0.198	0.482	1	0.488	0.871	0.59	1.286
	EMPLOYMENT(5)	-0.148	0.37	0.16	1	0.689	0.862	0.418	1.781
	GAD			68.104	3	0			
	GAD(1)	1.386	1.04	1.776	1	0.183	3.998	0.521	30.692
	GAD(2)	1.801	1.026	3.084	1	0.079	6.056	0.811	45.206
	GAD(3)	2.478	1.024	5.858	1	0.016	11.912	1.602	88.566
	DISCHARGE			22.649	5	0			
	DISCHARGE(1)	0.483	0.11	19.157	1	0	1.621	1.306	2.013
	DISCHARGE(2)	0.064	0.295	0.047	1	0.828	1.066	0.598	1.899
	DISCHARGE(3)	0.397	0.197	4.044	1	0.044	1.487	1.01	2.189
	DISCHARGE(4)	0.035	0.214	0.027	1	0.869	1.036	0.682	1.574
	DISCHARGE(5)	-0.112	0.256	0.192	1	0.661	0.894	0.541	1.477
	Constant	-3.094	1.037	8.894	1	0.003	0.045		

a. Variable(s) entered on step 1: MODEL, GENDER, DISABILITY, AGEREGRESSION, EMPLOYMENT, GAD, DISCHARGE.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 118: Logistical regression on no change outcome on PHQ moderate/severe for sensitivity analysis

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1a	MODEL(1)	-0.128	0.105	1.486	1	0.223	0.88	0.717	1.081
	GENDER(1)	0.037	0.108	0.116	1	0.733	1.037	0.84	1.282
	DISABILITY(1)	-0.216	0.204	1.115	1	0.291	0.806	0.54	1.203
	AGEREGRESSION			5.733	5	0.333			
	AGEREGRESSION(1)	-0.369	0.185	3.968	1	0.046	0.692	0.481	0.994
	AGEREGRESSION(2)	-0.261	0.194	1.812	1	0.178	0.77	0.527	1.126
	AGEREGRESSION(3)	-0.104	0.201	0.271	1	0.603	0.901	0.608	1.335
	AGEREGRESSION(4)	-0.223	0.226	0.974	1	0.324	0.8	0.514	1.246
	AGEREGRESSION(5)	-0.407	0.418	0.947	1	0.331	0.666	0.294	1.511
	EMPLOYMENT			25.876	5	0			
	EMPLOYMENT(1)	0.561	0.129	18.852	1	0	1.752	1.36	2.256
	EMPLOYMENT(2)	0.323	0.222	2.124	1	0.145	1.381	0.895	2.133
	EMPLOYMENT(3)	0.61	0.157	15.075	1	0	1.841	1.353	2.505
	EMPLOYMENT(4)	0.441	0.21	4.418	1	0.036	1.554	1.03	2.343
	EMPLOYMENT(5)	0.393	0.389	1.023	1	0.312	1.482	0.692	3.176
	GAD			6.904	3	0.075			
	GAD(1)	-0.836	0.519	2.588	1	0.108	0.434	0.157	1.2
	GAD(2)	-0.408	0.495	0.682	1	0.409	0.665	0.252	1.752
	GAD(3)	-0.599	0.492	1.483	1	0.223	0.549	0.209	1.441
	DISCHARGE			363.899	5	0			
	DISCHARGE(1)	2.064	0.121	289.094	1	0	7.877	6.209	9.993
	DISCHARGE(2)	2.175	0.263	68.194	1	0	8.8	5.252	14.746
	DISCHARGE(3)	1.916	0.192	100.045	1	0	6.794	4.668	9.89
	DISCHARGE(4)	2.17	0.196	122.11	1	0	8.755	5.959	12.865
	DISCHARGE(5)	1.527	0.236	41.992	1	0	4.605	2.901	7.308
	Constant	-1.6	0.526	9.263	1	0.002	0.202		

a. Variable(s) entered on step 1: MODEL, GENDER, DISABILITY, AGEREGRESSION, EMPLOYMENT, GAD, DISCHARGE.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 119: Logistical regression on reliably deteriorated outcome on PHQ moderate/severe for sensitivity analysis

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1a	MODEL(1)	-0.44	0.205	4.619	1	0.032	0.644	0.431	0.962
	GENDER(1)	-0.144	0.206	0.488	1	0.485	0.866	0.578	1.297
	DISABILITY(1)	0.385	0.329	1.366	1	0.243	1.47	0.771	2.803
	AGEREGRESSION			3.938	5	0.558			
	AGEREGRESSION(1)	-0.518	0.388	1.778	1	0.182	0.596	0.278	1.275
	AGEREGRESSION(2)	-0.068	0.382	0.032	1	0.858	0.934	0.441	1.976
	AGEREGRESSION(3)	-0.049	0.394	0.015	1	0.902	0.953	0.44	2.062
	AGEREGRESSION(4)	0.021	0.429	0.002	1	0.961	1.021	0.44	2.368
	AGEREGRESSION(5)	-0.292	0.72	0.165	1	0.685	0.746	0.182	3.061
	EMPLOYMENT			12.959	5	0.024			
	EMPLOYMENT(1)	0.689	0.252	7.458	1	0.006	1.991	1.215	3.265
	EMPLOYMENT(2)	-0.119	0.495	0.058	1	0.81	0.888	0.337	2.342
	EMPLOYMENT(3)	0.837	0.288	8.455	1	0.004	2.31	1.314	4.062
	EMPLOYMENT(4)	0.13	0.446	0.084	1	0.771	1.138	0.475	2.727
	EMPLOYMENT(5)	0.893	0.632	1.995	1	0.158	2.442	0.707	8.427
	GAD			54.907	3	0			
	GAD(1)	-0.646	0.66	0.957	1	0.328	0.524	0.144	1.912
	GAD(2)	-1.205	0.64	3.548	1	0.06	0.3	0.086	1.05
	GAD(3)	-2.474	0.649	14.512	1	0	0.084	0.024	0.301
	DISCHARGE			76.469	5	0			
	DISCHARGE(1)	1.782	0.261	46.548	1	0	5.943	3.562	9.916
	DISCHARGE(2)	2.197	0.418	27.583	1	0	8.994	3.962	20.414
	DISCHARGE(3)	1.882	0.357	27.731	1	0	6.565	3.259	13.225
	DISCHARGE(4)	2.425	0.326	55.443	1	0	11.3	5.969	21.394
	DISCHARGE(5)	1.224	0.518	5.575	1	0.018	3.401	1.231	9.394
	Constant	-2.446	0.744	10.807	1	0.001	0.087		

a. Variable(s) entered on step 1: MODEL, GENDER, DISABILITY, AGEREGRESSION, EMPLOYMENT, GAD, DISCHARGE.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 120: Logistical regression on recovered outcome on PHQ severe for sensitivity analysis

								Lower	Upper
Step 1a	MODEL(1)	0.474	0.105	20.472	1	0	1.606	1.308	1.972
	GENDER(1)	-0.068	0.106	0.406	1	0.524	0.935	0.759	1.15
	DISABILITY(1)	-0.051	0.178	0.082	1	0.774	0.95	0.671	1.346
	AGEREGRESSION			25.608	5	0			
	AGEREGRESSION(1)	-0.111	0.257	0.188	1	0.665	0.895	0.541	1.48
	AGEREGRESSION(2)	0.247	0.258	0.912	1	0.34	1.28	0.771	2.123
	AGEREGRESSION(3)	0.315	0.255	1.534	1	0.216	1.371	0.832	2.258
	AGEREGRESSION(4)	0.008	0.269	0.001	1	0.975	1.008	0.595	1.708
	AGEREGRESSION(5)	1.333	0.398	11.232	1	0.001	3.793	1.739	8.27
	EMPLOYMENT			102.04	5	0			
	EMPLOYMENT(1)	-1	0.129	60.477	1	0	0.368	0.286	0.473
	EMPLOYMENT(2)	-0.456	0.298	2.335	1	0.126	0.634	0.354	1.137
	EMPLOYMENT(3)	-1.258	0.153	67.466	1	0	0.284	0.211	0.384
	EMPLOYMENT(4)	-0.954	0.228	17.473	1	0	0.385	0.246	0.603
	EMPLOYMENT(5)	-1.193	0.326	13.374	1	0	0.303	0.16	0.575
	GAD			39.724	3	0			
	GAD(1)	0.633	0.782	0.656	1	0.418	1.883	0.407	8.713
	GAD(2)	0.132	0.733	0.033	1	0.857	1.142	0.271	4.803
	GAD(3)	-0.607	0.723	0.706	1	0.401	0.545	0.132	2.247
	DISCHARGE			456.046	5	0			
	DISCHARGE(1)	-3.102	0.187	274.258	1	0	0.045	0.031	0.065
	DISCHARGE(2)	-2.995	0.403	55.157	1	0	0.05	0.023	0.11
	DISCHARGE(3)	-2.378	0.239	98.817	1	0	0.093	0.058	0.148
	DISCHARGE(4)	-4.204	0.51	68.004	1	0	0.015	0.005	0.041
	DISCHARGE(5)	-1.412	0.254	30.992	1	0	0.244	0.148	0.401
	Constant	0.773	0.768	1.013	1	0.314	2.167		

a. Variable(s) entered on step 1: MODEL, GENDER, DISABILITY, AGEREGRESSION, EMPLOYMENT, GAD, DISCHARGE.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 121: Logistical regression on reliably improved outcome on PHQ severe for sensitivity analysis

								Lower	Upper
Step 1a	MODEL(1)	-0.054	0.082	0.431	1	0.512	0.948	0.808	1.112
	GENDER(1)	0.046	0.083	0.311	1	0.577	1.047	0.89	1.232
	DISABILITY(1)	-0.196	0.142	1.921	1	0.166	0.822	0.622	1.085
	AGEREGRESSION			20.973	5	0.001			
	AGEREGRESSION(1)	0.226	0.183	1.525	1	0.217	1.254	0.876	1.795
	AGEREGRESSION(2)	-0.072	0.185	0.151	1	0.697	0.931	0.648	1.337
	AGEREGRESSION(3)	-0.11	0.184	0.356	1	0.551	0.896	0.625	1.284
	AGEREGRESSION(4)	0.135	0.194	0.483	1	0.487	1.144	0.782	1.674
	AGEREGRESSION(5)	-0.844	0.339	6.187	1	0.013	0.43	0.221	0.836
	EMPLOYMENT			7.612	5	0.179			
	EMPLOYMENT(1)	0.11	0.1	1.203	1	0.273	1.116	0.917	1.359
	EMPLOYMENT(2)	0.22	0.23	0.912	1	0.34	1.246	0.793	1.956
	EMPLOYMENT(3)	0.208	0.116	3.199	1	0.074	1.231	0.98	1.547
	EMPLOYMENT(4)	0.172	0.177	0.945	1	0.331	1.188	0.839	1.682
	EMPLOYMENT(5)	0.614	0.271	5.115	1	0.024	1.848	1.085	3.146
	GAD			19.882	3	0			
	GAD(1)	-0.571	0.743	0.592	1	0.442	0.565	0.132	2.42
	GAD(2)	0.146	0.681	0.046	1	0.83	1.157	0.305	4.395
	GAD(3)	0.513	0.672	0.581	1	0.446	1.67	0.447	6.239
	DISCHARGE			15.748	5	0.008			
	DISCHARGE(1)	0.123	0.098	1.563	1	0.211	1.131	0.933	1.371
	DISCHARGE(2)	-0.255	0.202	1.594	1	0.207	0.775	0.522	1.151
	DISCHARGE(3)	0.238	0.151	2.502	1	0.114	1.269	0.945	1.705
	DISCHARGE(4)	-0.359	0.146	6.022	1	0.014	0.698	0.524	0.93
	DISCHARGE(5)	-0.254	0.216	1.391	1	0.238	0.775	0.508	1.183
	Constant	-1.135	0.699	2.636	1	0.104	0.322		

a. Variable(s) entered on step 1: MODEL, GENDER, DISABILITY, AGEREGRESSION, EMPLOYMENT, GAD, DISCHARGE.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 122: Logistical regression on no change outcome on PHQ severe for sensitivity analysis

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1a	MODEL(1)	-0.311	0.094	10.886	1	0.001	0.733	0.609	0.882
	GENDER(1)	0.04	0.095	0.174	1	0.677	1.041	0.863	1.255
	DISABILITY(1)	0.324	0.155	4.376	1	0.036	1.383	1.021	1.873
	AGEREGRESSION			2.844	5	0.724			
	AGEREGRESSION(1)	-0.319	0.204	2.453	1	0.117	0.727	0.487	1.084
	AGEREGRESSION(2)	-0.206	0.203	1.031	1	0.31	0.814	0.546	1.212
	AGEREGRESSION(3)	-0.18	0.202	0.792	1	0.374	0.835	0.562	1.241
	AGEREGRESSION(4)	-0.167	0.216	0.598	1	0.439	0.846	0.555	1.292
	AGEREGRESSION(5)	-0.246	0.377	0.426	1	0.514	0.782	0.373	1.637
	EMPLOYMENT			43.71	5	0			
	EMPLOYMENT(1)	0.666	0.115	33.514	1	0	1.946	1.554	2.439
	EMPLOYMENT(2)	0.252	0.264	0.908	1	0.341	1.287	0.766	2.16
	EMPLOYMENT(3)	0.71	0.134	28.224	1	0	2.034	1.565	2.644
	EMPLOYMENT(4)	0.54	0.205	6.969	1	0.008	1.716	1.149	2.563
	EMPLOYMENT(5)	0.257	0.323	0.633	1	0.426	1.293	0.687	2.433
	GAD			31.205	3	0			
	GAD(1)	-0.566	0.819	0.477	1	0.49	0.568	0.114	2.829
	GAD(2)	-0.491	0.749	0.43	1	0.512	0.612	0.141	2.655
	GAD(3)	0.282	0.737	0.147	1	0.702	1.326	0.313	5.618
	DISCHARGE			452.678	5	0			
	DISCHARGE(1)	2.085	0.116	323.148	1	0	8.042	6.407	10.094
	DISCHARGE(2)	2.187	0.199	120.18	1	0	8.904	6.023	13.163
	DISCHARGE(3)	1.85	0.162	129.81	1	0	6.358	4.625	8.741
	DISCHARGE(4)	2.453	0.153	255.784	1	0	11.619	8.603	15.694
	DISCHARGE(5)	1.864	0.213	76.61	1	0	6.45	4.249	9.792
	Constant	-2.253	0.77	8.57	1	0.003	0.105		

a. Variable(s) entered on step 1: MODEL, GENDER, DISABILITY, AGEREGRESSION, EMPLOYMENT, GAD, DISCHARGE.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 123: Logistical regression on reliably deteriorated outcome on PHQ severe for sensitivity analysis

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1a	MODEL(1)	-0.196	0.282	0.485	1	0.486	0.822	0.473	1.427
	GENDER(1)	-0.494	0.301	2.687	1	0.101	0.61	0.338	1.101
	DISABILITY(1)	0.103	0.408	0.064	1	0.8	1.109	0.498	2.469
	AGEREGRESSION			3.989	5	0.551			
	AGEREGRESSION(1)	1.364	0.862	2.502	1	0.114	3.911	0.722	21.198
	AGEREGRESSION(2)	1.577	0.872	3.27	1	0.071	4.842	0.876	26.765
	AGEREGRESSION(3)	1.241	0.881	1.985	1	0.159	3.459	0.616	19.436
	AGEREGRESSION(4)	1.244	0.908	1.877	1	0.171	3.469	0.585	20.565
	AGEREGRESSION(5)	0.601	1.652	0.132	1	0.716	1.823	0.072	46.443
	EMPLOYMENT			10.133	5	0.072			
	EMPLOYMENT(1)	1.021	0.383	7.108	1	0.008	2.776	1.311	5.881
	EMPLOYMENT(2)	0.432	0.955	0.205	1	0.651	1.541	0.237	10.007
	EMPLOYMENT(3)	1.183	0.429	7.612	1	0.006	3.264	1.409	7.563
	EMPLOYMENT(4)	1.151	0.557	4.274	1	0.039	3.162	1.062	9.416
	EMPLOYMENT(5)	0.261	1.442	0.033	1	0.856	1.298	0.077	21.937
	GAD			95.173	3	0			
	GAD(1)	0.494	1.233	0.161	1	0.688	1.64	0.146	18.387
	GAD(2)	-0.123	1.172	0.011	1	0.917	0.885	0.089	8.797
	GAD(3)	-2.81	1.175	5.723	1	0.017	0.06	0.006	0.602
	DISCHARGE			41.927	5	0			
	DISCHARGE(1)	2.237	0.449	24.804	1	0	9.367	3.884	22.593
	DISCHARGE(2)	3.029	0.545	30.927	1	0	20.678	7.11	60.137
	DISCHARGE(3)	1.798	0.588	9.341	1	0.002	6.036	1.906	19.116
	DISCHARGE(4)	2.84	0.495	32.976	1	0	17.11	6.491	45.1
	DISCHARGE(5)	1.149	1.106	1.08	1	0.299	3.155	0.361	27.567
	Constant	-5.504	1.514	13.215	1	0	0.004		

a. Variable(s) entered on step 1: MODEL, GENDER, DISABILITY, AGEREGRESSION, EMPLOYMENT, GAD, DISCHARGE.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 124: Logistical regression on recovered outcome on GAD moderate for sensitivity analysis

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	MODEL(1)	0.4	0.099	16.165	1	0	1.491	1.227	1.812
	GENDER(1)	0.146	0.104	1.977	1	0.16	1.157	0.944	1.418
	DISABILITY(1)	0.086	0.205	0.177	1	0.674	1.09	0.73	1.629
	AGEREGRESSION			16.742	5	0.005			
	AGEREGRESSION(1)	0.319	0.193	2.729	1	0.099	1.376	0.942	2.008
	AGEREGRESSION(2)	0.444	0.204	4.719	1	0.03	1.559	1.044	2.328
	AGEREGRESSION(3)	0.509	0.208	5.971	1	0.015	1.664	1.106	2.503
	AGEREGRESSION(4)	0.275	0.222	1.531	1	0.216	1.316	0.852	2.033
	AGEREGRESSION(5)	1.395	0.392	12.665	1	0	4.036	1.872	8.705
	EMPLOYMENT			60.493	5	0			
	EMPLOYMENT(1)	-0.89	0.138	41.855	1	0	0.411	0.314	0.538
	EMPLOYMENT(2)	-0.3	0.213	1.989	1	0.158	0.741	0.488	1.124
	EMPLOYMENT(3)	-0.93	0.171	29.481	1	0	0.395	0.282	0.552
	EMPLOYMENT(4)	-0.477	0.208	5.242	1	0.022	0.621	0.413	0.934
	EMPLOYMENT(5)	-0.823	0.335	6.047	1	0.014	0.439	0.228	0.846
	PHQ			49.312	4	0			
	PHQ(1)	-0.006	0.319	0	1	0.985	0.994	0.532	1.858
	PHQ(2)	-0.369	0.304	1.469	1	0.225	0.692	0.381	1.256
	PHQ(3)	-0.895	0.304	8.68	1	0.003	0.408	0.225	0.741
	PHQ(4)	-0.893	0.32	7.796	1	0.005	0.409	0.219	0.766
	DISCHARGE			471.923	5	0			
	DISCHARGE(1)	-2.674	0.141	361.144	1	0	0.069	0.052	0.091
	DISCHARGE(2)	-2.044	0.324	39.825	1	0	0.13	0.069	0.244
	DISCHARGE(3)	-1.76	0.2	77.646	1	0	0.172	0.116	0.254
	DISCHARGE(4)	-3.51	0.434	65.292	1	0	0.03	0.013	0.07
	DISCHARGE(5)	-1.029	0.26	15.703	1	0	0.358	0.215	0.595
	Constant	1.237	0.342	13.052	1	0	3.445		

a. Variable(s) entered on step 1: MODEL, GENDER, DISABILITY, AGEREGRESSION, EMPLOYMENT, PHQ, DISCHARGE.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 125: Logistical regression on reliably improved outcome on GAD moderate for sensitivity analysis

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	MODEL(1)	-0.25	0.115	4.739	1	0.029	0.778	0.621	0.975
	GENDER(1)	-0.14	0.121	1.34	1	0.247	0.869	0.685	1.102
	DISABILITY(1)	0.004	0.235	0	1	0.987	1.004	0.634	1.59
	AGEREGRESSION			7.335	5	0.197			
	AGEREGRESSION(1)	0.198	0.218	0.825	1	0.364	1.219	0.795	1.871
	AGEREGRESSION(2)	-0.071	0.234	0.093	1	0.76	0.931	0.589	1.473
	AGEREGRESSION(3)	-0.03	0.238	0.016	1	0.901	0.971	0.609	1.546
	AGEREGRESSION(4)	0.057	0.257	0.049	1	0.824	1.059	0.64	1.751
	AGEREGRESSION(5)	-0.883	0.519	2.894	1	0.089	0.414	0.15	1.144
	EMPLOYMENT			1.624	5	0.898			
	EMPLOYMENT(1)	0.114	0.153	0.556	1	0.456	1.121	0.831	1.512
	EMPLOYMENT(2)	0.03	0.245	0.015	1	0.903	1.03	0.638	1.665
	EMPLOYMENT(3)	-0.037	0.198	0.036	1	0.85	0.963	0.654	1.42
	EMPLOYMENT(4)	0.136	0.234	0.336	1	0.562	1.146	0.724	1.813
	EMPLOYMENT(5)	0.379	0.441	0.741	1	0.389	1.461	0.616	3.465
	PHQ			89.179	4	0			
	PHQ(1)	1.18	0.742	2.531	1	0.112	3.254	0.761	13.919
	PHQ(2)	1.481	0.727	4.151	1	0.042	4.398	1.058	18.284
	PHQ(3)	2.268	0.724	9.824	1	0.002	9.661	2.339	39.902
	PHQ(4)	2.716	0.729	13.888	1	0	15.116	3.623	63.058
	DISCHARGE			22.741	5	0			
	DISCHARGE(1)	0.449	0.136	10.861	1	0.001	1.567	1.2	2.046
	DISCHARGE(2)	-0.628	0.447	1.979	1	0.16	0.534	0.222	1.28
	DISCHARGE(3)	0.572	0.218	6.905	1	0.009	1.772	1.156	2.714
	DISCHARGE(4)	0.53	0.265	3.997	1	0.046	1.699	1.01	2.857
	DISCHARGE(5)	-0.459	0.413	1.232	1	0.267	0.632	0.281	1.421
	Constant	-3.661	0.747	24.002	1	0	0.026		

a. Variable(s) entered on step 1: MODEL, GENDER, DISABILITY, AGEREGRESSION, EMPLOYMENT, PHQ, DISCHARGE.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 126: Logistical regression on no change outcome on GAD moderate for sensitivity analysis

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1a	MODEL(1)	-0.258	0.106	5.887	1	0.015	0.773	0.628	0.952
	GENDER(1)	0.09	0.11	0.678	1	0.41	1.095	0.883	1.357
	DISABILITY(1)	0.021	0.216	0.009	1	0.924	1.021	0.668	1.56
	AGEREGRESSION			9.125	5	0.104			
	AGEREGRESSION(1)	-0.488	0.189	6.691	1	0.01	0.614	0.424	0.889
	AGEREGRESSION(2)	-0.439	0.2	4.836	1	0.028	0.645	0.436	0.953
	AGEREGRESSION(3)	-0.505	0.207	5.981	1	0.014	0.603	0.403	0.905
	AGEREGRESSION(4)	-0.375	0.226	2.748	1	0.097	0.688	0.441	1.071
	AGEREGRESSION(5)	-0.917	0.459	3.992	1	0.046	0.4	0.163	0.983
	EMPLOYMENT			15.576	5	0.008			
	EMPLOYMENT(1)	0.421	0.139	9.124	1	0.003	1.524	1.159	2.003
	EMPLOYMENT(2)	0.226	0.213	1.124	1	0.289	1.254	0.825	1.906
	EMPLOYMENT(3)	0.568	0.176	10.467	1	0.001	1.766	1.251	2.491
	EMPLOYMENT(4)	0.293	0.219	1.789	1	0.181	1.34	0.873	2.057
	EMPLOYMENT(5)	0.255	0.413	0.382	1	0.536	1.291	0.575	2.898
	PHQ			21.21	4	0			
	PHQ(1)	-0.097	0.374	0.068	1	0.795	0.907	0.436	1.888
	PHQ(2)	0.393	0.355	1.225	1	0.268	1.481	0.739	2.97
	PHQ(3)	0.331	0.356	0.865	1	0.352	1.392	0.693	2.796
	PHQ(4)	-0.235	0.375	0.394	1	0.53	0.79	0.379	1.648
	DISCHARGE			267.156	5	0			
	DISCHARGE(1)	1.893	0.122	242.251	1	0	6.637	5.229	8.423
	DISCHARGE(2)	1.248	0.288	18.816	1	0	3.482	1.981	6.118
	DISCHARGE(3)	1.424	0.197	52.247	1	0	4.153	2.823	6.11
	DISCHARGE(4)	1.726	0.236	53.544	1	0	5.616	3.538	8.916
	DISCHARGE(5)	1.303	0.274	22.681	1	0	3.679	2.152	6.29
	Constant	-1.881	0.387	23.615	1	0	0.152		

a. Variable(s) entered on step 1: MODEL, GENDER, DISABILITY, AGEREGRESSION, EMPLOYMENT, PHQ, DISCHARGE.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 127: Logistical regression on reliably deteriorated outcome on GAD moderate for sensitivity analysis

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	MODEL(1)	0.049	0.18	0.075	1	0.784	1.051	0.738	1.496
	GENDER(1)	-0.472	0.194	5.921	1	0.015	0.624	0.427	0.912
	DISABILITY(1)	-0.069	0.33	0.043	1	0.835	0.933	0.489	1.783
	AGEREGRESSION			2.035	5	0.844			
	AGEREGRESSION(1)	0.075	0.32	0.054	1	0.816	1.077	0.576	2.016
	AGEREGRESSION(2)	0.178	0.338	0.276	1	0.599	1.194	0.616	2.317
	AGEREGRESSION(3)	0.114	0.352	0.105	1	0.746	1.121	0.562	2.233
	AGEREGRESSION(4)	0.374	0.378	0.979	1	0.323	1.454	0.693	3.053
	AGEREGRESSION(5)	-0.272	0.737	0.136	1	0.713	0.762	0.18	3.234
	EMPLOYMENT			40.242	5	0			
	EMPLOYMENT(1)	1.3	0.235	30.669	1	0	3.668	2.316	5.811
	EMPLOYMENT(2)	0.591	0.39	2.299	1	0.129	1.806	0.841	3.878
	EMPLOYMENT(3)	1.432	0.276	26.841	1	0	4.189	2.436	7.201
	EMPLOYMENT(4)	0.623	0.361	2.972	1	0.085	1.864	0.918	3.783
	EMPLOYMENT(5)	1.36	0.619	4.819	1	0.028	3.896	1.157	13.119
	PHQ			11.447	4	0.022			
	PHQ(1)	-0.654	0.517	1.596	1	0.207	0.52	0.189	1.434
	PHQ(2)	-1.303	0.504	6.682	1	0.01	0.272	0.101	0.73
	PHQ(3)	-1.236	0.5	6.107	1	0.013	0.291	0.109	0.774
	PHQ(4)	-1.226	0.522	5.51	1	0.019	0.293	0.105	0.817
	DISCHARGE			135.053	5	0			
	DISCHARGE(1)	2.31	0.239	93.639	1	0	10.079	6.312	16.094
	DISCHARGE(2)	3.276	0.347	88.968	1	0	26.482	13.405	52.315
	DISCHARGE(3)	1.763	0.35	25.433	1	0	5.832	2.939	11.574
	DISCHARGE(4)	2.905	0.339	73.476	1	0	18.259	9.398	35.474
	DISCHARGE(5)	1.795	0.48	13.986	1	0	6.017	2.349	15.409
	Constant	-3.639	0.577	39.746	1	0	0.026		

a. Variable(s) entered on step 1: MODEL, GENDER, DISABILITY, AGEREGRESSION, EMPLOYMENT, PHQ, DISCHARGE.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 128: Logistical regression on recovered outcome on GAD7 severe for sensitivity analysis

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1a	MODEL(1)	0.436	0.076	33.002	1	0	1.546	1.333	1.794
	GENDER(1)	-0.084	0.078	1.161	1	0.281	0.919	0.788	1.072
	DISABILITY(1)	-0.064	0.144	0.2	1	0.655	0.938	0.707	1.244
	AGEREGRESSION			14.996	5	0.01			
	AGEREGRESSION(1)	-0.011	0.16	0.005	1	0.946	0.989	0.723	1.353
	AGEREGRESSION(2)	0.078	0.164	0.225	1	0.635	1.081	0.784	1.49
	AGEREGRESSION(3)	0.104	0.166	0.392	1	0.531	1.11	0.801	1.537
	AGEREGRESSION(4)	-0.009	0.178	0.003	1	0.959	0.991	0.699	1.404
	AGEREGRESSION(5)	0.91	0.286	10.158	1	0.001	2.485	1.42	4.349
	EMPLOYMENT			99.486	5	0			
	EMPLOYMENT(1)	-0.761	0.096	63.186	1	0	0.467	0.387	0.564
	EMPLOYMENT(2)	-0.26	0.189	1.898	1	0.168	0.771	0.532	1.116
	EMPLOYMENT(3)	-0.919	0.119	59.525	1	0	0.399	0.316	0.504
	EMPLOYMENT(4)	-0.526	0.161	10.645	1	0.001	0.591	0.431	0.811
	EMPLOYMENT(5)	-0.819	0.249	10.788	1	0.001	0.441	0.27	0.719
	PHQ			73.543	4	0			
	PHQ(1)	-0.016	0.42	0.001	1	0.969	0.984	0.432	2.239
	PHQ(2)	-0.584	0.391	2.233	1	0.135	0.558	0.259	1.2
	PHQ(3)	-0.799	0.384	4.318	1	0.038	0.45	0.212	0.956
	PHQ(4)	-1.222	0.384	10.133	1	0.001	0.295	0.139	0.625
	DISCHARGE			815.696	5	0			
	DISCHARGE(1)	-3.044	0.134	518.957	1	0	0.048	0.037	0.062
	DISCHARGE(2)	-2.593	0.309	70.371	1	0	0.075	0.041	0.137
	DISCHARGE(3)	-2.526	0.201	157.837	1	0	0.08	0.054	0.119
	DISCHARGE(4)	-3.382	0.299	128.223	1	0	0.034	0.019	0.061
	DISCHARGE(5)	-1.312	0.183	51.392	1	0	0.269	0.188	0.385
	Constant	1.328	0.412	10.376	1	0.001	3.774		

a. Variable(s) entered on step 1: MODEL, GENDER, DISABILITY, AGEREGRESSION, EMPLOYMENT, PHQ, DISCHARGE.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 129: Logistical regression on reliable improvement outcome on GAD7 severe for sensitivity analysis

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1a	MODEL(1)	-0.101	0.064	2.504	1	0.114	0.904	0.798	1.024
	GENDER(1)	0.071	0.065	1.176	1	0.278	1.074	0.944	1.221
	DISABILITY(1)	-0.033	0.12	0.077	1	0.781	0.967	0.764	1.224
	AGEREGRESSION			15.81	5	0.007			
	AGEREGRESSION(1)	0.115	0.129	0.794	1	0.373	1.121	0.872	1.443
	AGEREGRESSION(2)	-0.097	0.132	0.537	1	0.464	0.908	0.7	1.176
	AGEREGRESSION(3)	-0.089	0.134	0.439	1	0.507	0.915	0.703	1.191
	AGEREGRESSION(4)	0.081	0.144	0.318	1	0.573	1.085	0.818	1.439
	AGEREGRESSION(5)	-0.581	0.256	5.127	1	0.024	0.56	0.338	0.925
	EMPLOYMENT			3.965	5	0.554			
	EMPLOYMENT(1)	0.062	0.079	0.609	1	0.435	1.064	0.911	1.243
	EMPLOYMENT(2)	0.143	0.16	0.799	1	0.371	1.153	0.844	1.577
	EMPLOYMENT(3)	0.109	0.096	1.274	1	0.259	1.115	0.923	1.346
	EMPLOYMENT(4)	-0.036	0.138	0.067	1	0.795	0.965	0.737	1.264
	EMPLOYMENT(5)	0.32	0.217	2.175	1	0.14	1.377	0.9	2.107
	PHQ			35.131	4	0			
	PHQ(1)	0.392	0.45	0.76	1	0.383	1.48	0.613	3.574
	PHQ(2)	0.776	0.421	3.394	1	0.065	2.173	0.952	4.963
	PHQ(3)	0.997	0.416	5.742	1	0.017	2.711	1.199	6.128
	PHQ(4)	1.168	0.415	7.903	1	0.005	3.215	1.424	7.257
	DISCHARGE			32.798	5	0			
	DISCHARGE(1)	0.346	0.076	20.68	1	0	1.414	1.218	1.641
	DISCHARGE(2)	0.002	0.176	0	1	0.991	1.002	0.71	1.415
	DISCHARGE(3)	0.281	0.126	4.952	1	0.026	1.324	1.034	1.696
	DISCHARGE(4)	-0.216	0.127	2.907	1	0.088	0.806	0.629	1.033
	DISCHARGE(5)	-0.123	0.169	0.531	1	0.466	0.884	0.635	1.231
	Constant	-1.796	0.433	17.214	1	0	0.166		

a. Variable(s) entered on step 1: MODEL, GENDER, DISABILITY, AGEREGRESSION, EMPLOYMENT, PHQ, DISCHARGE.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 130: Logistical regression on no change outcome on GAD7 severe for sensitivity analysis

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	MODEL(1)	-0.234	0.075	9.672	1	0.002	0.791	0.683	0.917
	GENDER(1)	0.002	0.078	0.001	1	0.977	1.002	0.861	1.167
	DISABILITY(1)	0.094	0.137	0.465	1	0.495	1.098	0.839	1.438
	AGEREGRESSION			2.375	5	0.795			
	AGEREGRESSION(1)	-0.05	0.148	0.112	1	0.738	0.951	0.711	1.273
	AGEREGRESSION(2)	0.027	0.152	0.032	1	0.859	1.027	0.763	1.383
	AGEREGRESSION(3)	0.058	0.155	0.141	1	0.708	1.06	0.783	1.435
	AGEREGRESSION(4)	-0.051	0.168	0.091	1	0.763	0.951	0.684	1.321
	AGEREGRESSION(5)	-0.279	0.304	0.843	1	0.359	0.756	0.417	1.373
	EMPLOYMENT			61.077	5	0			
	EMPLOYMENT(1)	0.577	0.092	39.141	1	0	1.78	1.486	2.133
	EMPLOYMENT(2)	0.202	0.194	1.092	1	0.296	1.224	0.838	1.789
	EMPLOYMENT(3)	0.712	0.111	41.154	1	0	2.038	1.639	2.533
	EMPLOYMENT(4)	0.593	0.155	14.607	1	0	1.809	1.335	2.452
	EMPLOYMENT(5)	0.566	0.26	4.738	1	0.03	1.761	1.058	2.931
	PHQ			32.612	4	0			
	PHQ(1)	-0.326	0.49	0.445	1	0.505	0.722	0.276	1.883
	PHQ(2)	-0.108	0.442	0.06	1	0.806	0.897	0.378	2.132
	PHQ(3)	0.145	0.433	0.113	1	0.737	1.156	0.495	2.702
	PHQ(4)	0.448	0.431	1.081	1	0.298	1.566	0.672	3.645
	DISCHARGE			722.887	5	0			
	DISCHARGE(1)	2.105	0.09	544.448	1	0	8.204	6.875	9.79
	DISCHARGE(2)	2.13	0.176	146.545	1	0	8.418	5.962	11.885
	DISCHARGE(3)	1.937	0.134	209.925	1	0	6.935	5.337	9.012
	DISCHARGE(4)	2.391	0.129	342.927	1	0	10.93	8.486	14.078
	DISCHARGE(5)	1.887	0.167	128.165	1	0	6.601	4.761	9.152
	Constant	-2.599	0.456	32.542	1	0	0.074		

a. Variable(s) entered on step 1: MODEL, GENDER, DISABILITY, AGEREGRESSION, EMPLOYMENT, PHQ, DISCHARGE.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 131: Logistical regression on reliable deterioration outcome on GAD7 severe for sensitivity analysis

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	MODEL(1)	-0.532	0.218	5.944	1	0.015	0.587	0.383	0.901
	GENDER(1)	-0.257	0.228	1.269	1	0.26	0.773	0.494	1.21
	DISABILITY(1)	0.471	0.385	1.496	1	0.221	1.601	0.753	3.403
	AGEREGRESSION			11.585	5	0.041			
	AGEREGRESSION(1)	-0.674	0.418	2.608	1	0.106	0.509	0.225	1.155
	AGEREGRESSION(2)	0.224	0.388	0.335	1	0.563	1.251	0.585	2.675
	AGEREGRESSION(3)	-0.35	0.428	0.669	1	0.414	0.704	0.304	1.631
	AGEREGRESSION(4)	0.083	0.44	0.036	1	0.85	1.087	0.458	2.576
	AGEREGRESSION(5)	0.692	0.809	0.732	1	0.392	1.998	0.409	9.76
	EMPLOYMENT			16.916	5	0.005			
	EMPLOYMENT(1)	0.813	0.254	10.232	1	0.001	2.256	1.37	3.713
	EMPLOYMENT(2)	-0.813	0.777	1.095	1	0.295	0.443	0.097	2.035
	EMPLOYMENT(3)	0.411	0.343	1.439	1	0.23	1.509	0.77	2.955
	EMPLOYMENT(4)	0.479	0.417	1.318	1	0.251	1.614	0.713	3.657
	EMPLOYMENT(5)	-1.894	1.172	2.61	1	0.106	0.151	0.015	1.497
	PHQ			75.71	4	0			
	PHQ(1)	-0.729	0.792	0.847	1	0.357	0.482	0.102	2.279
	PHQ(2)	-0.259	0.667	0.151	1	0.698	0.772	0.209	2.853
	PHQ(3)	-1.406	0.669	4.417	1	0.036	0.245	0.066	0.91
	PHQ(4)	-2.759	0.687	16.14	1	0	0.063	0.016	0.243
	DISCHARGE			66.925	5	0			
	DISCHARGE(1)	1.868	0.293	40.525	1	0	6.477	3.644	11.513
	DISCHARGE(2)	2.204	0.499	19.546	1	0	9.065	3.412	24.089
	DISCHARGE(3)	2.264	0.364	38.65	1	0	9.623	4.713	19.647
	DISCHARGE(4)	2.559	0.349	53.869	1	0	12.926	6.526	25.601
	DISCHARGE(5)	-15.917	2794.991	0	1	0.995	0	0	.
	Constant	-3.439	0.782	19.354	1	0	0.032		

a. Variable(s) entered on step 1: MODEL, GENDER, DISABILITY, AGEREGRESSION, EMPLOYMENT, PHQ, DISCHARGE.

Key of Variable Reference Categories

	Gender	Disability	Age	Employment	PHQ	GAD	Discharge
Reference	Female	No	18-24	Employed	Minimal	Minimal	Completed
1	Male	Yes	25-34	Unemployed	Mild	Mild	Dropped Out
2			35-44	Student	Moderate	Moderate	Not Suitable
3			45-54	SickDisabled	Moderate to Severe	Severe	Declined Treatment
4			55-64	Homemaker	Severe		Referred On
5			65-74	Retired			

Appendix 132: Cross tabulation on drop out and recovery outcome for whole service

Discharge Reason				Model		Total
				Allocated	Progressive	
Not Suitable	Recovery	Recovered	Count	25	37	62
			% within Model	12.20%	15.80%	14.10%
		Reliable Improvement	Count	48	67	115
			% within Model	23.40%	28.60%	26.20%
		Non Recovered	Count	94	108	202
			% within Model	45.90%	46.20%	46.00%
		Reliable Deterioration	Count	38	22	60
			% within Model	18.50%	9.40%	13.70%
	Total		Count	205	234	439
			% within Model	100.00%	100.00%	100.00%
Declined Treatment	Recovery	Recovered	Count	86	106	192
			% within Model	22.80%	16.90%	19.10%
		Reliable Improvement	Count	100	182	282
			% within Model	26.50%	28.90%	28.00%
		Non Recovered	Count	155	295	450
			% within Model	41.00%	46.90%	44.70%
		Reliable Deterioration	Count	37	46	83
			% within Model	9.80%	7.30%	8.20%
	Total		Count	378	629	1007
			% within Model	100.00%	100.00%	100.00%
Completed Treatment	Recovery	Recovered	Count	1945	3964	5909
			% within Model	61.90%	67.10%	65.30%
		Reliable Improvement	Count	708	1247	1955
			% within Model	22.50%	21.10%	21.60%
		Non Recovered	Count	428	614	1042
			% within Model	13.60%	10.40%	11.50%
		Reliable Deterioration	Count	62	86	148
			% within Model	2.00%	1.50%	1.60%
	Total		Count	3143	5911	9054
			% within Model	100.00%	100.00%	100.00%
Dropped Out of Treatment	Recovery	Recovered	Count	210	218	428
			% within Model	15.20%	11.10%	12.80%
		Reliable Improvement	Count	420	635	1055
			% within Model	30.40%	32.40%	31.60%
		Non Recovered	Count	640	933	1573
			% within Model	46.30%	47.60%	47.10%
		Reliable Deterioration	Count	112	174	286
			% within Model	8.10%	8.90%	8.60%
	Total		Count	1382	1960	3342
			% within Model	100.00%	100.00%	100.00%
Referred On	Recovery	Recovered	Count	27	34	61
			% within Model	7.40%	5.20%	6.00%
		Reliable Improvement	Count	91	159	250
			% within Model	24.80%	24.40%	24.50%
		Non Recovered	Count	195	378	573
			% within Model	53.10%	58.00%	56.20%
		Reliable Deterioration	Count	54	81	135
			% within Model	14.70%	12.40%	13.20%
	Total		Count	367	652	1019
			% within Model	100.00%	100.00%	100.00%
Total	Recovery	Recovered	Count	2293	4359	6652
			% within Model	41.90%	46.40%	44.80%
		Reliable Improvement	Count	1367	2290	3657
			% within Model	25.00%	24.40%	24.60%
		Non Recovered	Count	1512	2328	3840
			% within Model	27.60%	24.80%	25.80%
		Reliable Deterioration	Count	303	409	712
			% within Model	5.50%	4.40%	4.80%
	Total		Count	5475	9386	14861
			% within Model	100.00%	100.00%	100.00%

Appendix 133: Chi square on drop out and recovery outcome for whole service

Discharge Reason		Value	df	Asymp. Sig. (2-sided)
Not Suitable	Pearson Chi-Square	8.821[b]	3	0.032
	Likelihood Ratio	8.864	3	0.031
	Linear-by-Linear Association	6.372	1	0.012
	N of Valid Cases	439		
Declined Treatment	Pearson Chi-Square	8.419[c]	3	0.038
	Likelihood Ratio	8.316	3	0.04
	Linear-by-Linear Association	1.398	1	0.237
	N of Valid Cases	1007		
Completed Treatment	Pearson Chi-Square	32.340[d]	3	0
	Likelihood Ratio	31.846	3	0
	Linear-by-Linear Association	31.643	1	0
	N of Valid Cases	9054		
Dropped Out of Treatment	Pearson Chi-Square	12.387[e]	3	0.006
	Likelihood Ratio	12.248	3	0.007
	Linear-by-Linear Association	5.717	1	0.017
	N of Valid Cases	3342		
Referred On	Pearson Chi-Square	3.725[f]	3	0.293
	Likelihood Ratio	3.671	3	0.299
	Linear-by-Linear Association	0.241	1	0.624
	N of Valid Cases	1019		
Total	Pearson Chi-Square	37.109[a]	3	0
	Likelihood Ratio	36.978	3	0
	Linear-by-Linear Association	36.804	1	0
	N of Valid Cases	14861		

Appendix 134: Cross tabulation on drop out and recovery outcome for north cohort

Discharge				Model		Total
				Allocated	Progressive	
Not Suitable	Recovery	Recovered	Count	11	21	32
			% within Model	13.10%	14.00%	13.70%
		Reliable Improvement	Count	20	44	64
			% within Model	23.80%	29.30%	27.40%
		Non Recovered	Count	36	69	105
			% within Model	42.90%	46.00%	44.90%
		Reliable Deterioration	Count	17	16	33
			% within Model	20.20%	10.70%	14.10%
	Total		Count	84	150	234
			% within Model	100.00%	100.00%	100.00%
Declined	Recovery	Recovered	Count	51	57	108
			% within Model	24.40%	17.30%	20.10%
		Reliable Improvement	Count	58	89	147
			% within Model	27.80%	27.10%	27.30%
		Non Recovered	Count	84	153	237
			% within Model	40.20%	46.50%	44.10%
		Reliable Deterioration	Count	16	30	46
			% within Model	7.70%	9.10%	8.60%
	Total		Count	209	329	538
			% within Model	100.00%	100.00%	100.00%
Completed	Recovery	Recovered	Count	1176	2312	3488
			% within Model	64.40%	68.70%	67.20%
		Reliable Improvement	Count	396	680	1076
			% within Model	21.70%	20.20%	20.70%
		Non Recovered	Count	227	321	548
			% within Model	12.40%	9.50%	10.60%
		Reliable Deterioration	Count	28	50	78
			% within Model	1.50%	1.50%	1.50%
	Total		Count	1827	3363	5190
			% within Model	100.00%	100.00%	100.00%
Dropped	Recovery	Recovered	Count	118	98	216
			% within Model	13.80%	8.80%	11.00%
		Reliable Improvement	Count	259	354	613
			% within Model	30.30%	31.90%	31.20%
		Non Recovered	Count	404	553	957
			% within Model	47.30%	49.80%	48.70%
		Reliable Deterioration	Count	73	106	179
			% within Model	8.50%	9.50%	9.10%
	Total		Count	854	1111	1965
			% within Model	100.00%	100.00%	100.00%
Referred On	Recovery	Recovered	Count	14	16	30
			% within Model	7.10%	4.50%	5.40%
		Reliable Improvement	Count	45	97	142
			% within Model	23.00%	27.10%	25.60%
		Non Recovered	Count	106	199	305
			% within Model	54.10%	55.60%	55.10%
		Reliable Deterioration	Count	31	46	77
			% within Model	15.80%	12.80%	13.90%
	Total		Count	196	358	554
			% within Model	100.00%	100.00%	100.00%
Total	Recovery	Recovered	Count	1370	2504	3874
			% within Model	43.20%	47.10%	45.70%
		Reliable Improvement	Count	778	1264	2042
			% within Model	24.50%	23.80%	24.10%
		Non Recovered	Count	857	1295	2152
			% within Model	27.00%	24.40%	25.40%
		Reliable Deterioration	Count	165	248	413
			% within Model	5.20%	4.70%	4.90%
	Total		Count	3170	5311	8481
			% within Model	100.00%	100.00%	100.00%

Appendix 135: Chi square on drop out and recovery outcome for north cohort

Discharge Reason		Value	df	Asymp. Sig. (2-sided)
Not Suitable	Pearson Chi-Square	4.249b	3	0.236
	Likelihood Ratio	4.111	3	0.25
	Linear-by-Linear Association	1.923	1	0.166
	N of Valid Cases	234		
Declined Treatment	Pearson Chi-Square	4.688c	3	0.196
	Likelihood Ratio	4.64	3	0.2
	Linear-by-Linear Association	4.169	1	0.041
	N of Valid Cases	538		
Completed Treatment	Pearson Chi-Square	13.903d	3	0.003
	Likelihood Ratio	13.701	3	0.003
	Linear-by-Linear Association	11.631	1	0.001
	N of Valid Cases	5190		
Dropped Out of Treatment	Pearson Chi-Square	12.457e	3	0.006
	Likelihood Ratio	12.339	3	0.006
	Linear-by-Linear Association	6.646	1	0.01
	N of Valid Cases	1965		
Referred On	Pearson Chi-Square	3.371f	3	0.338
	Likelihood Ratio	3.31	3	0.346
	Linear-by-Linear Association	0.069	1	0.792
	N of Valid Cases	554		
Total	Pearson Chi-Square	13.835a	3	0.003
	Likelihood Ratio	13.829	3	0.003
	Linear-by-Linear Association	12.998	1	0
	N of Valid Cases	8481		

Appendix 136: Cross tabulation on drop out and recovery outcome for south cohort

Discharge Reason				Model		Total
				Allocated	Progressive	
Not Suitable	Recovery	Recovered	Count	22	8	30
			% within Model	14.50%	17.00%	15.10%
		Reliable Improvement	Count	36	13	49
			% within Model	23.70%	27.70%	24.60%
		Non Recovered	Count	70	23	93
			% within Model	46.10%	48.90%	46.70%
		Reliable Deterioration	Count	24	3	27
			% within Model	15.80%	6.40%	13.60%
	Total		Count	152	47	199
			% within Model	100.00%	100.00%	100.00%
Declined Treatment	Recovery	Recovered	Count	58	26	84
			% within Model	18.20%	18.20%	18.20%
		Reliable Improvement	Count	83	52	135
			% within Model	26.00%	36.40%	29.20%
		Non Recovered	Count	150	57	207
			% within Model	47.00%	39.90%	44.80%
		Reliable Deterioration	Count	28	8	36
			% within Model	8.80%	5.60%	7.80%
	Total		Count	319	143	462
			% within Model	100.00%	100.00%	100.00%
Completed Treatment	Recovery	Recovered	Count	1467	915	2382
			% within Model	58.20%	71.40%	62.70%
		Reliable Improvement	Count	619	245	864
			% within Model	24.60%	19.10%	22.70%
		Non Recovered	Count	374	111	485
			% within Model	14.80%	8.70%	12.80%
		Reliable Deterioration	Count	59	11	70
			% within Model	2.30%	0.90%	1.80%
	Total		Count	2519	1282	3801
			% within Model	100.00%	100.00%	100.00%
Dropped Out of Treatment	Recovery	Recovered	Count	156	53	209
			% within Model	16.80%	12.40%	15.40%
		Reliable Improvement	Count	298	140	438
			% within Model	32.20%	32.80%	32.40%
		Non Recovered	Count	405	198	603
			% within Model	43.70%	46.40%	44.60%
		Reliable Deterioration	Count	67	36	103
			% within Model	7.20%	8.40%	7.60%
	Total		Count	926	427	1353
			% within Model	100.00%	100.00%	100.00%
Referred On	Recovery	Recovered	Count	23	8	31
			% within Model	6.80%	6.70%	6.80%
		Reliable Improvement	Count	79	26	105
			% within Model	23.20%	21.80%	22.90%
		Non Recovered	Count	195	70	265
			% within Model	57.40%	58.80%	57.70%
		Reliable Deterioration	Count	43	15	58
			% within Model	12.60%	12.60%	12.60%
	Total		Count	340	119	459
			% within Model	100.00%	100.00%	100.00%
Total	Recovery	Recovered	Count	1726	1010	2736
			% within Model	40.60%	50.00%	43.60%
		Reliable Improvement	Count	1115	476	1591
			% within Model	26.20%	23.60%	25.40%
		Non Recovered	Count	1194	459	1653
			% within Model	28.10%	22.70%	26.30%
		Reliable Deterioration	Count	221	73	294
			% within Model	5.20%	3.60%	4.70%
	Total		Count	4256	2018	6274
			% within Model	100.00%	100.00%	100.00%

Appendix 137: Chi square on drop out and recovery outcome for south cohort

Discharge Reason		Value	df	Asymp. Sig. (2-sided)
Not Suitable	Pearson Chi-Square	2.790b	3	0.425
	Likelihood Ratio	3.192	3	0.363
	Linear-by-Linear Association	1.496	1	0.221
	N of Valid Cases	199		
Declined Treatment	Pearson Chi-Square	6.030c	3	0.11
	Likelihood Ratio	5.998	3	0.112
	Linear-by-Linear Association	2.362	1	0.124
	N of Valid Cases	462		
Completed Treatment	Pearson Chi-Square	70.210d	3	0
	Likelihood Ratio	73.207	3	0
	Linear-by-Linear Association	68.718	1	0
	N of Valid Cases	3801		
Dropped Out of Treatment	Pearson Chi-Square	4.757e	3	0.19
	Likelihood Ratio	4.895	3	0.18
	Linear-by-Linear Association	3.69	1	0.055
	N of Valid Cases	1353		
Referred On	Pearson Chi-Square	.107f	3	0.991
	Likelihood Ratio	0.108	3	0.991
	Linear-by-Linear Association	0.032	1	0.859
	N of Valid Cases	459		
Total	Pearson Chi-Square	53.874a	3	0
	Likelihood Ratio	53.992	3	0
	Linear-by-Linear Association	50.096	1	0
	N of Valid Cases	6274		

Appendix 138: Cross tabulation on drop out and recovery outcome for sensitivity analysis

Discharge				Embedded		Total
				Year 2	Year 4	
Not Suitable	Recovery	Recovered	Count	12	22	34
			% within	9.20%	18.50%	13.70%
		Reliable Improvement	Count	30	31	61
			% within	23.10%	26.10%	24.50%
		Non Recovered	Count	71	56	127
			% within	54.60%	47.10%	51.00%
		Reliable Deterioration	Count	17	10	27
			% within	13.10%	8.40%	10.80%
	Total		Count	130	119	249
			% within	100.00%	100.00%	100.00%
Declined	Recovery	Recovered	Count	51	42	93
			% within	22.00%	13.20%	16.90%
		Reliable Improvement	Count	61	95	156
			% within	26.30%	30.00%	28.40%
		Non Recovered	Count	97	154	251
			% within	41.80%	48.60%	45.70%
		Reliable Deterioration	Count	23	26	49
			% within	9.90%	8.20%	8.90%
	Total		Count	232	317	549
			% within	100.00%	100.00%	100.00%
Completed	Recovery	Recovered	Count	1328	2032	3360
			% within	58.70%	71.80%	66.00%
		Reliable Improvement	Count	552	524	1076
			% within	24.40%	18.50%	21.10%
		Non Recovered	Count	336	246	582
			% within	14.90%	8.70%	11.40%
		Reliable Deterioration	Count	45	30	75
			% within	2.00%	1.10%	1.50%
	Total		Count	2261	2832	5093
			% within	100.00%	100.00%	100.00%
Dropped	Recovery	Recovered	Count	124	104	228
			% within	13.90%	10.80%	12.30%
		Reliable Improvement	Count	286	317	603
			% within	32.20%	32.90%	32.50%
		Non Recovered	Count	408	457	865
			% within	45.90%	47.40%	46.70%
		Reliable Deterioration	Count	71	86	157
			% within	8.00%	8.90%	8.50%
	Total		Count	889	964	1853
			% within	100.00%	100.00%	100.00%
Referred On	Recovery	Recovered	Count	16	16	32
			% within	5.50%	6.30%	5.90%
		Reliable Improvement	Count	72	71	143
			% within	24.60%	28.10%	26.20%
		Non Recovered	Count	168	134	302
			% within	57.30%	53.00%	55.30%
		Reliable Deterioration	Count	37	32	69
			% within	12.60%	12.60%	12.60%
	Total		Count	293	253	546
			% within	100.00%	100.00%	100.00%
Total	Recovery	Recovered	Count	1531	2216	3747
			% within	40.20%	49.40%	45.20%
		Reliable Improvement	Count	1001	1038	2039
			% within	26.30%	23.10%	24.60%
		Non Recovered	Count	1080	1047	2127
			% within	28.40%	23.30%	25.70%
		Reliable Deterioration	Count	193	184	377
			% within	5.10%	4.10%	4.50%
	Total		Count	3805	4485	8290
			% within	100.00%	100.00%	100.00%

Appendix 139: Chi square on drop out and recovery outcome for sensitivity analysis

Discharge Reason		Value	df	Asymp. Sig. (2-sided)
Not Suitable	Pearson Chi-Square	6.070b	3	0.108
	Likelihood Ratio	6.127	3	0.106
	Linear-by-Linear Association	5.787	1	0.016
	N of Valid Cases	249		
Declined Treatment	Pearson Chi-Square	8.451c	3	0.038
	Likelihood Ratio	8.371	3	0.039
	Linear-by-Linear Association	2.547	1	0.111
	N of Valid Cases	549		
Completed Treatment	Pearson Chi-Square	102.421d	3	0
	Likelihood Ratio	102.174	3	0
	Linear-by-Linear Association	98.317	1	0
	N of Valid Cases	5093		
Dropped Out of Treatment	Pearson Chi-Square	4.529e	3	0.21
	Likelihood Ratio	4.527	3	0.21
	Linear-by-Linear Association	2.973	1	0.085
	N of Valid Cases	1853		
Referred On	Pearson Chi-Square	1.274f	3	0.735
	Likelihood Ratio	1.273	3	0.736
	Linear-by-Linear Association	0.655	1	0.418
	N of Valid Cases	546		
Total	Pearson Chi-Square	71.327a	3	0
	Likelihood Ratio	71.491	3	0
	Linear-by-Linear Association	61.046	1	0
	N of Valid Cases	8290		