An evaluation of a public online intervention for depression

Kylie Bennett

June 2019

A thesis submitted for the degree of Doctor of Philosophy of

The Australian National University

Statement of originality

"I, Kylie Bennett, declare that the PhD thesis entitled 'An evaluation of a public online intervention for depression' is no more than 100,000 words in length including quotes and exclusive of tables, figures, appendices, bibliography, references and footnotes. This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work."

Kylie Bennett
June 2019

Acknowledgements

I would like to thank my supervisory panel, including Professors Britt Klein and John Cunningham, for their guidance and supervision. Most particularly, I extend my gratitude to Emeritus Professor Kathy Griffiths, the chair of my panel, for her guidance, mentorship and above all for her unwavering belief that I could complete this thesis while working full-time. Kathy is also the co-creator of the *MoodGYM* program, which is evaluated in this thesis.

I would also like to thank Anthony Bennett not only for his support at a personal level but also for his IT expertise. Anthony has managed the technical implementation of the *MoodGYM* program and has enabled the online trials described in this thesis. He has facilitated many PhD projects at the ANU Centre for Mental Health Research, and I am grateful that this has included my project.

The observational trials described in Chapters 4 and 5 relied on the public delivery of e-hub Web Services by the Australian National University, which was funded by the Commonwealth Department of Health. e-hub Web Services were overseen by the e-hub Management Team of which I was a member, along with Professor Kathy Griffiths, Professor Helen Christensen, Mr Anthony Bennett and Ms Julia Reynolds. The trial described in Chapter 6 was designed in collaboration with Professors Helen Christensen, Kathy Griffiths and Andrew Mackinnon. The RCT described in Chapter 7 was designed in collaboration with Professors John Powell, Kathy Griffiths and Helen Christensen. This trial was funded by NHS Choices, and I managed the trial in collaboration with a team at the University of Warwick.

Abstract

Depression is a leading cause of disability worldwide, but only a minority of individuals receive evidence-based treatment for the condition. Moreover, there are barriers to accessing effective therapies such as cognitive behaviour therapy (CBT). Internet-delivered CBT (iCBT) is a scalable, accessible and potentially effective alternative to face-to-face CBT. This thesis is a detailed study of the unguided use among spontaneous and other community users of 'MoodGYM', a well-known, open-access, automated iCBT program for depression. The thesis comprises two systematic reviews of the effectiveness of iCBT and four empirical studies of *MoodGYM*, including two observational studies, a partially randomised controlled trial (PRPT) and a randomised controlled trial (RCT). The first review – a synthesis of previous systematic reviews – concluded that iCBT for depression is broadly effective for the treatment of depression. The second review focused specifically on automated interventions which were entirely self-directed and involved no direct human input in their delivery. It found some evidence for the effectiveness of automated iCBT, particularly for adults in community settings. However, findings from controlled efficacy trials may not translate into equivalent user outcomes in real-world settings, and there is a lack of systematically documented information about the user characteristics and outcomes of globally accessible online programs for mental health. Accordingly, the first observational study (N = 417,354) investigated the characteristics of spontaneous, community users of automated *MoodGYM* and found they were mostly female, highly educated and had high initial depression scores. Nearly half were referred by a health professional, with male users and those from the U.K. or rural areas more likely to be referred. The second observational study (N = 417,354) examined adherence and outcomes and their predictors among spontaneous users. Overall there was a small but significant decrease in depression symptoms among users (Cohen's d = 0.28). Completion of modules was best predicted by initial motivation, and improved depression by higher adherence and baseline depression and lower anxiety. The third empirical study (PRPT; N = 23,988) investigated differences in outcomes for participants who chose a program variant compared to those who agreed to randomisation, and which modules were associated with the best outcomes. Older, more educated or more anxious users were more likely to choose their program and choosing was a significant predictor of both adherence and depressive symptom improvement, although the latter was not significant after controlling for the program variant received. The introductory and cognitive restructuring modules (Modules 1 and 2) were the most critical program components. Finally, an RCT (N = 3,070) of unguided *MoodGYM* users examined the potential predictors and moderators of treatment effect. Although several variables predicted depression outcomes across all participants, the only outcome moderator (associated with the intervention but not the control) was self-reported ability to wash or

dress oneself. Open-access *MoodGYM* has been accessed by a large number of public users, with small but significant reductions in depression symptoms, suggesting that automated iCBT has the potential for a significant public health impact. However, increasing user engagement may improve depression outcomes. The current studies suggest that motivating and educating users prior to intervention use and providing a choice of program components may increase adherence. There may also be value in emphasising the first two *MoodGYM* modules. This thesis provides some insight into for whom *MoodGYM* is most effective; however, further research is required, particularly into moderators of treatment effect and the attitudes and expectations of sub-groups of users to maximise the public health impact of automated iCBT.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Background

There is a high prevalence of depressive disorders worldwide. In 2015, over 300 million individuals were estimated to suffer from depression (approximately 4.4% of the global population) [1]. Depression is a major contributor to the global burden of disease [2] and the leading cause of disability worldwide, with over 50 million Years Lived with Disability (YLDs) estimated in 2015 [1].

There are effective treatments for depression. However, the majority of individuals with depression do not receive evidence-based treatment (e.g., [3]). Barriers to receiving treatment include cost, insufficient access to services, lack of mental health literacy, stigma and a preference for self-reliance rather than professional help [4].

One effective treatment for depression is cognitive behaviour therapy (CBT) [5] – a psychological therapy traditionally delivered face-to-face by an appropriately trained psychologist or other mental health professional. Technology advances over the last 30 years, including the proliferation of personal computers and more recently, the widespread use of the Internet, have increasingly been incorporated into models of CBT delivery [6, 7, 8]. Such innovation potentially removes barriers to accessing treatment – for example, CBT delivered via video-conferencing removes access barriers related to distance [9], CBT training programs on stand-alone computers decreases the impact of a lack of trained therapists [10], and online CBT training programs can potentially circumvent the barrier of stigma associated with mental health problems if they can be accessed anonymously [11].

In particular, Internet-delivered CBT (iCBT) programs for depression which are fully automated (that is, do not employ any direct human input in their delivery) offer enormous potential at a global level since they are highly scalable. Once developed, such interventions can be delivered to large numbers of individuals at a low incremental cost [12]. Combined with the global reach of the Internet, this creates an opportunity to provide accessible, anonymous treatment resources to the millions of individuals worldwide with depression and depressive symptoms. If such programs are shown to be effective for the treatment of depression, the provision of wide-spread or open-access iCBT could produce a significant public health impact.

The present thesis is a detailed study of community users of 'MoodGYM' – a well-known, open-access, automated iCBT program for depression.

1.2 Definitions

A number of terms associated with Internet-based cognitive behaviour therapy are used throughout the thesis. In the interests of clarity and to avoid ambiguity, these terms, as they are employed in this thesis, are defined briefly below.

Internet-delivered CBT (iCBT): An intervention which predominantly teaches CBT techniques, and which is delivered through the Internet, accessible as a website through a web browser. CBT delivered through a smartphone app is not classified as iCBT within this thesis.

Computerised CBT (cCBT): An intervention which predominantly teaches CBT techniques, and which is delivered on a stand-alone computer, for example through a CD-ROM.

Guided intervention: An intervention which includes human guidance as a therapeutic component. For example, guidance could be provided by a therapist, health professional, lay person or coach, and might consist of phone calls, web chat or emails on either a regular basis or to provide feedback on completed activities.

Unguided or self-guided intervention: An intervention which does not include any human guidance as a therapeutic component. The provision of self-guided interventions may include human support (or contact from researchers in research settings) which is not directly related to the intervention content.

Automated intervention: A type of unguided intervention which does not include any human guidance or direct human input for its delivery. The provision of automated interventions in research settings does not include any human-initiated contact. For example, a trial protocol which involves weekly phone calls from a research assistant in addition to use of an unguided intervention would not be classified as an automated intervention. Open-access intervention: An intervention which is publicly available to all individuals with Internet access. Users of an open-access intervention can access the intervention at any time, from any location.

Spontaneous, community users: Individuals who access open-access interventions outside of research contexts.

Adherence: Engagement with an intervention, for example, by logging in to the intervention and through completion of modules or exercises.

Drop-out: When a research participant fails to complete a post-intervention or follow-up outcome measure. Note that drop-out does not refer to (lack of) user adherence to an intervention.

Completer: A research participant who completes all outcome measures that are specified in a research protocol. Note that completer does not refer to user adherence to the intervention.

1.3 The MoodGYM iCBT intervention

The empirical studies in this thesis focus solely on the online program *MoodGYM*. One of the first automated iCBT programs to be delivered as an open-access intervention, *MoodGYM* was developed by researchers at the Australian National University and launched in 2001. To access the program, spontaneous, community users are required to agree to the terms of use and register with a username and password which they can use to access the program at any time. At registration users also complete a brief survey which includes demographic and other questions.

A detailed description of *MoodGYM* can be found in the MoodGYM Clinician's Manual [13]. In summary, *MoodGYM* is structured as five modules which provide CBT training through character stories, quizzes, interactive exercises and diaries. The program content specifically addresses depression and does not focus on comorbid conditions such as anxiety or insomnia. The first 'Feelings module' provides a brief introduction to the fundamentals of cognitive therapy. The second 'Thoughts module' extends this training to introduce dysfunctional thinking – 'warpy thoughts' – and provides users with examples and exercises to practice identifying their dysfunctional thoughts. The 'Unwarping module' then presents cognitive restructuring strategies for users to modify or 'unwarp' these thoughts and provides behavioural activation training. The fourth 'Destressing module' focuses on stressful events and relaxation methods, and the final 'Relationships module' delivers problem-solving training and examines how to cope with relationship break-ups. The content is reinforced by a cast of fictional characters who demonstrate the links between thoughts, feelings and behaviours. Figure 1.1 (a) - (e) displays screenshots from Version 3 of the publicly available program which was freely available worldwide from 2009 to 2017.

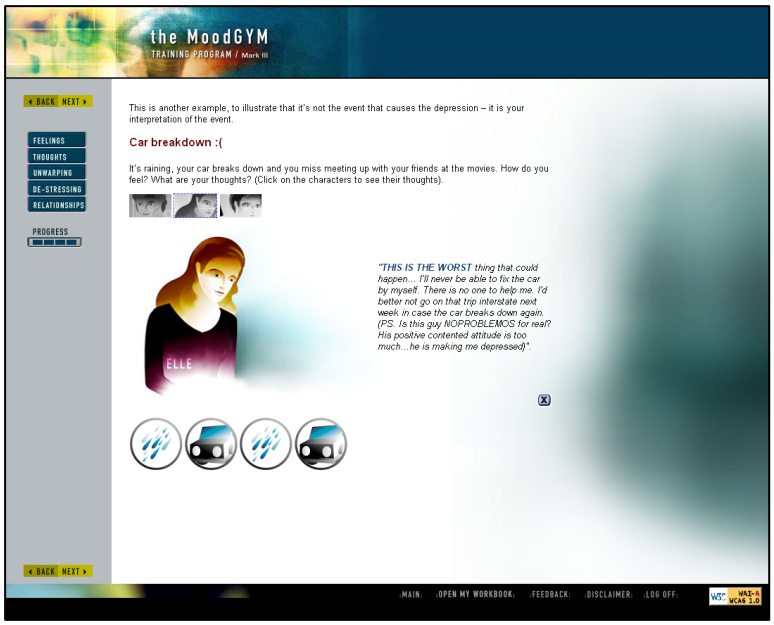

Figure 1.1 (a). Module 1 – Feelings. Introducing the concepts of CBT.

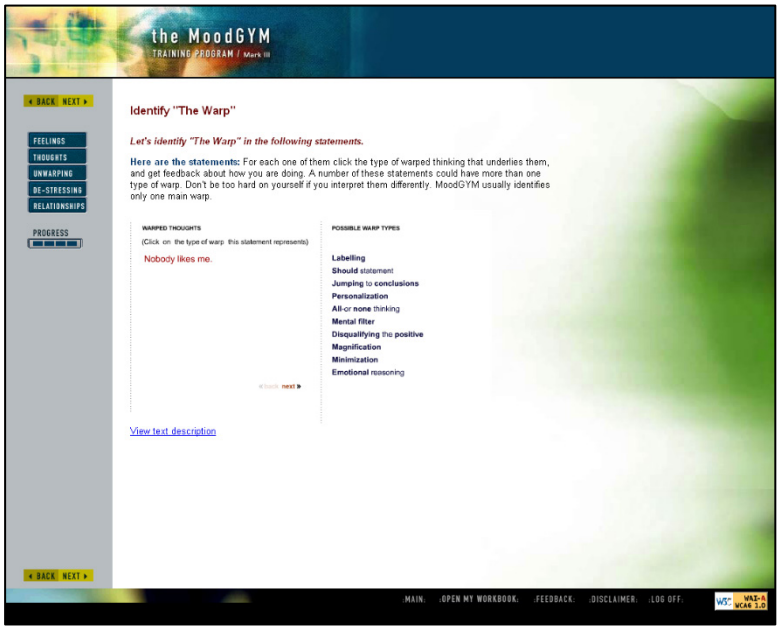

Figure 1.1 (b). Module 2 - Thoughts. Training in the identification of dysfunctional (warpy) thoughts.

Figure 1.1 (c). Module 3 - Unwarping. Cognitive restructuring (challenging warpy thoughts) and behavioural activation (pleasant events scheduling).

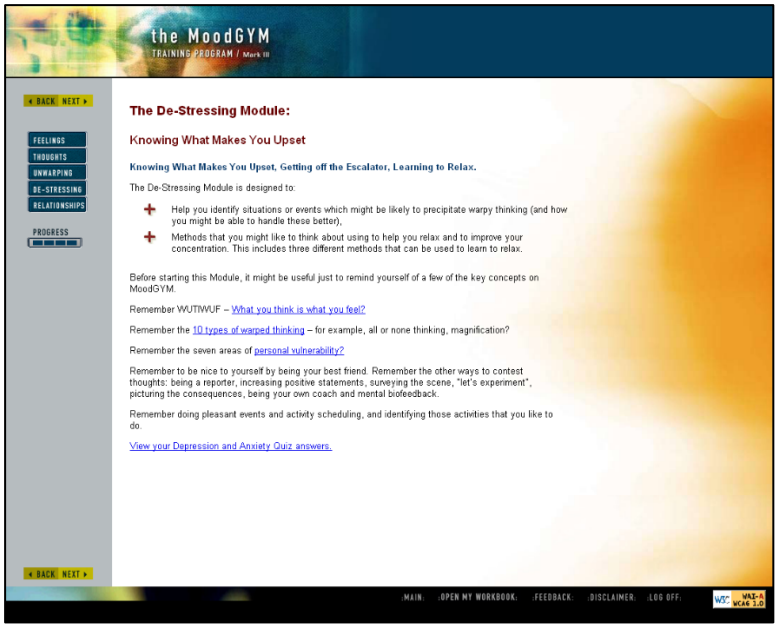

Figure 1.1 (d). Module 4 - Destressing. Stressful events and relaxation training.

Figure 1.1 (e). Module 5 - Relationships. Problem-solving training and coping with relationship issues.

The *MoodGYM* modules must be completed in linear order for the user to progress through the program. Modules 2 and 3 rely on the content taught and build on the responses provided by the user in the previous modules; however, Modules 4 and 5 do not. The program makes extensive use of interactive exercises in which the user practices identifying, categorising and reframing dysfunctional thoughts. As part of their use of the program users are encouraged to make text entries in workbooks and diaries, and the program also includes a workbook to review exercises and quizzes or to add entries to diaries. The program does not include any video. Use of the program is completely self-directed and there are no automated emails or other prompting messages such as text messages.

At the beginning of each module, users must complete a Depression and Anxiety symptom scale and are immediately provided with feedback about their level of symptoms, based on normative community data. Users with high numbers of depression or anxiety symptoms are encouraged to seek professional face-to-face help, and all users are provided with information about different sources of help. However, program access is not restricted based on user symptom levels – all users can continue to access the content regardless of their level of symptoms. Figure 1.2 shows a screenshot of the Depression Quiz, and Figure 1.3 shows results and feedback from this assessment.

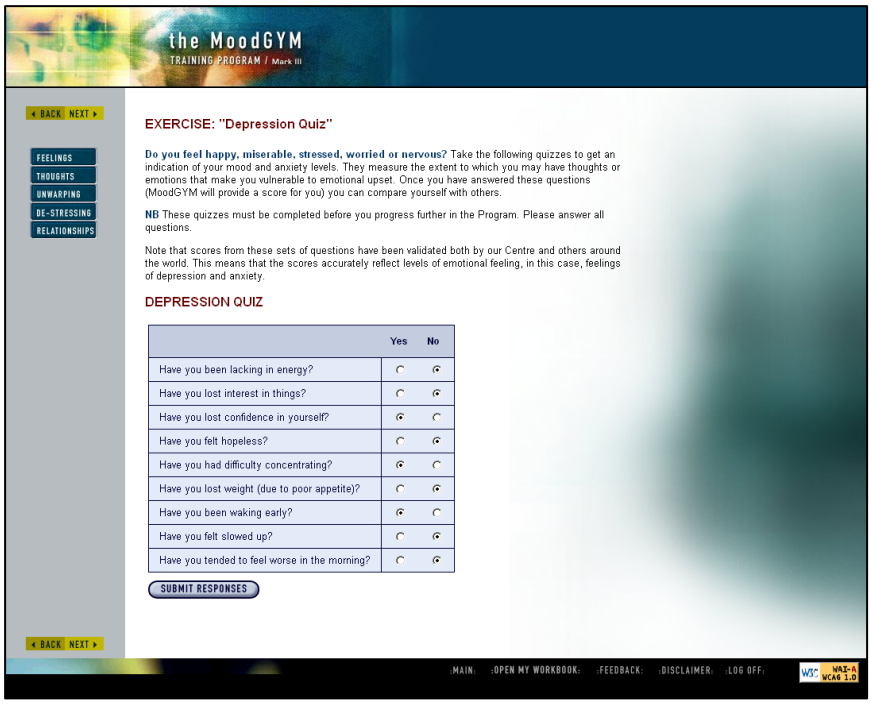

Figure 1.2. Screenshot from the MoodGYM program (Version 3) - Depression Quiz.

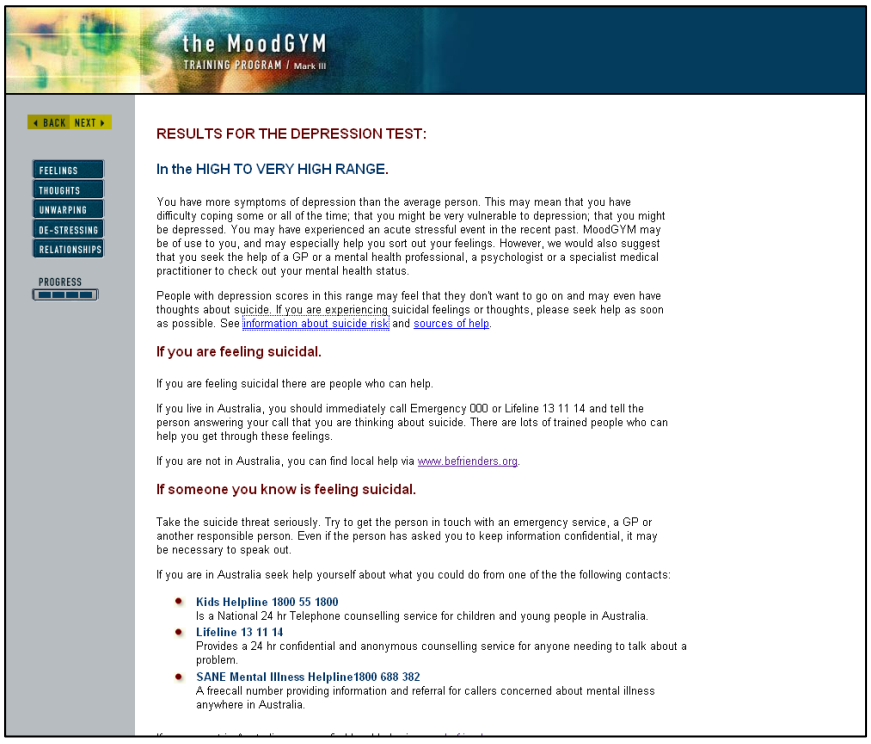

Figure 1.3. Screenshot from the MoodGYM program (Version 3) - Depression assessment results and feedback.

Directing users to other services does raise the possibility that any improvement in *MoodGYM* user symptoms is attributable to the help-seeking behaviour that the feedback might generate. Research on the impact of online depression interventions on help-seeking behaviour is sparse. One small randomised controlled trial (RCT) found that a brief, online intervention providing normative feedback and sources of help for depression and anxiety *did not* significantly affect help-seeking behaviour in a sample of elite athletes [14]. Likewise, a recent systematic review of interventions to promote help-seeking for mental health problems found that there was no evidence from four RCTs that online interventions were effective for increasing help-seeking behaviours in users [15]. However, notwithstanding the potential barriers to accessing other services which may have brought the user to *MoodGYM*, the possibility that users access alternative resources on the recommendation of the intervention cannot be excluded, particularly in observational studies of real-world program users.

MoodGYM was first released for public use in April 2001. Version 2 of the program, released in September 2003, delivered technical improvements and bug fixes to the original program. This version also introduced the requirement for users to complete core depression and anxiety assessments before progressing through the program. In March 2009 Version 3 was released on a new software platform, which enabled larger numbers of concurrent users and enhanced capabilities for research and translation. The content of this version was identical to the previous versions and translations for Norwegian, Finnish, German, Chinese and Dutch languages were also added. The program was implemented with a traditional website design and was not responsive for display on devices with smaller screens (such as smartphones)[1].

1.4 The current thesis

1.4.1 Research aims

The aim of the current thesis is to understand the potential public health impact of an automated, open-access iCBT program for depression by examining the unguided use of the *MoodGYM* program by community users, including user characteristics, program usage and depression outcomes. Additionally, the thesis considers how this impact might be maximised by asking: *Which modules are*

[1] The program's user experience was significantly updated when the Version 4 was released in July 2017. This most recent version includes technology improvements to ensure compatibility with modern delivery standards and accessibility requirements, and a responsive web design so the program can be used on devices with different sized screens. However, this version was not used in any of the studies presented in this thesis.

most important for depression improvement?; *Does user preference impact outcomes?*; and *Who benefits most from automated iCBT?*

1.4.2 Outline and contributions of the current thesis

The current thesis comprises eight chapters. Following the current brief Introduction (Chapter 1), Chapters 2 and 3 explore the extant literature on the effectiveness of iCBT for the treatment or prevention of depression. Chapter 2 is a systematic meta-review of existing systematic reviews of randomised controlled trials (RCTs) that have examined the effectiveness of iCBT for depression. The review extends previous meta-reviews by considering the effectiveness of both guided and unguided iCBT interventions: including the use of iCBT for the *prevention* as well as the *treatment* of depression; not restricting the target population groups; and including reviews which focus on participant adherence to iCBT and the cost-effectiveness of iCBT for depression. Chapter 3 then specifically considers the evidence for the effectiveness of automated iCBT interventions. Although previous systematic reviews have considered unguided iCBT more generally (including non-automated interventions) or, in addition have also included trials of non-iCBT interventions, to our knowledge, this is the first systematic review to focus specifically on RCTs of automated iCBT for depression.

The thesis then explores the characteristics and depression outcomes of 'real-world' spontaneous, community users of the open-access *MoodGYM* program. In the largest and most comprehensive study of its type to date, Chapters 4 and 5 examine a sample of over 400,000 spontaneous users who registered on the publicly available website over a period of approximately 5 years from March 2009. The observational study described in Chapter 4 examines the characteristics of these users and compares demographic characteristics and user pathways to the program, based on the country of registration. The study also investigates whether demographic sub-groups of these spontaneous users are more likely to have been referred to the program by a health professional or to have higher baseline depression symptoms. Chapter 5 then investigates the usage patterns and depression outcomes for these spontaneous users, and in particular, whether improved depression outcomes can be predicted by demographic or other variables. These variables include the user's level of motivation and their understanding of CBT – potentially modifiable factors which have not previously been examined in the context of outcomes for spontaneous users.

Given the low reported adherence rates in users of self-guided iCBT, Chapter 6 investigates depression outcomes and program adherence for users accessing different combinations of *MoodGYM* modules, to explore whether there are specific program components that are most

critical for improving symptoms in users. For this study (*N* = 23,988), spontaneous, community users were recruited to a Partially Randomised Preference Trial (PRPT), which allowed users to either choose or be randomised to one of several versions of *MoodGYM* that differ in module composition (e.g., the first module only vs the first and second modules combined). The design of this study also allows the first analysis of the role of preference on user adherence and depression outcomes by examining whether users who choose their intervention version are more likely to complete their program and/or have superior depression outcomes to users who are randomised. In addition, the demographic and clinical characteristics of users who choose (rather than agree to be randomised) are examined.

Chapter 7 presents a secondary analysis of data from an RCT of automated *MoodGYM* (*N* = 3,070), which recruited participants via the United Kingdom National Health Services consumer health information portal, NHS Choices, and in which participants were randomised to either receive *MoodGYM* or were placed in a wait-list control condition. The RCT design enables potential moderators of iCBT treatment to be explored – that is, whether there are demographic or baseline clinical characteristics of participants which affect depression outcomes for participants in the iCBT condition but not in the control condition. This study extends the limited existing research into moderators of iCBT treatment, to help understand which users might benefit most from automated iCBT. Although the study was undertaken in a research trial context (which is necessary for moderator research since a control condition is required for comparison), the setting approximates that of spontaneous, community users. In order to consider the data in the context of the earlier chapters in the thesis, predictors of depression improvement (across both conditions) are also investigated.

The thesis concludes in Chapter 8 with a summary of the findings and their implications for public health and closes with future research directions.

2. Systematic Review of iCBT for depression

2.1 Introduction

The first randomised controlled trial (RCT) of an iCBT intervention for depression, which evaluated the effectiveness of the 'Overcoming Depression on the Internet' (ODIN) program, was published by Clarke et al. in 2002 [16]. Although this initial study did not find an effect for the intervention compared to a control condition, there has since been a rapid growth of RCTs evaluating iCBT for depression. There has also been a corresponding growth in the publication of systematic reviews of RCTs of iCBT for depression since the first such reviews were published in 2006 [17, 18].

Despite the large number of systematic reviews of depression iCBT, to date, there have only been two published reviews of these reviews [19, 20]. In 2011, Foroushani and colleagues conducted a meta-review of systematic reviews focusing on the efficacy of computerised CBT (including iCBT) for depression in adults (with or without anxiety) [19]. The aim of this review was to examine the efficacy of iCBT in the treatment of depression, with an emphasis on the available evidence for individual programs. Twelves reviews were included in the analysis, which incorporated an evaluation of five specific programs. The authors 'cautiously' concluded that certain cCBT packages, specifically 'MoodGYM', 'Beating the Blues' and 'Colour Your Life', can have a positive effect on symptoms of depression but emphasised that the evidence was 'limited' [19]. The review also noted that cost-effectiveness information was only available for the *Beating the Blues* intervention, from one published paper included in one of the reviews.

A more recent, but more general review examined meta-analyses of Internet-delivered health interventions, including but not limited to iCBT for depression [20]. The aim of this 2017 review was to identify publicly available and effective self-help interventions through a search of meta-analyses of RCTs of interventions delivered over the Internet that included at least one intervention not guided by a clinician/therapist and which reported any health-related outcome. Overall, 71 meta-analyses were identified, comprising 268 unique RCTs of unguided Internet interventions, of which 15 RCTs evaluated depression iCBT programs. Of the depression interventions evaluated in these RCTs, four were publicly available (*MoodGYM*, 'BluePages', 'Deprexis', *Colour Your Life*). The review concluded that all four websites were associated with a reduction in depression symptoms. Three of the websites offered iCBT, whereas *BluePages* provided psycho-education.

However, there are limitations associated with the existing meta-reviews. In order to investigate the potential public health impact of iCBT for depression, it is necessary to consider not only the

effectiveness of interventions for the treatment of depression but also their effectiveness for depression prevention. Evidence needs to be assembled not only for adult populations but also for youth at a time of the first emergence of depression and a critical period for preventive intervention. The delivery model of different interventions is also pertinent – particularly whether evaluated programs are guided by a clinician, lay person or other coach, or are self-guided. This is especially important given that different models are associated with different resource implications and accessibility considerations. The extent to which iCBT program users engage with interventions (user adherence) and whether the delivery of programs is cost-effective are also highly relevant. Finally, a consideration of the quality of the evidence is critical when undertaking a review.

A number of the above factors were not considered by one or both of the extant meta-reviews. The very broad 2017 review [20] covered all health conditions but contained little discussion of depression specifically. Its emphasis was on dissemination and the availability of websites for which evidence was available, rather than on the quality of this evidence. The review also focused on self-help programs rather than iCBT specifically and did not examine user adherence or cost-effectiveness.

The Foroushani review [19] was focused on reviews of cCBT for depression. However, the review was limited to adult populations and treatment interventions. Study quality and bias were examined, as was cost-effectiveness; however, the review did not consider user adherence. Moreover, the review is no longer current, given the ongoing exponential growth in research involving iCBT interventions.

To address the limitation of the existing reviews, the current study is a systematic meta-review of all previous systematic reviews involving RCTs of the effectiveness of iCBT for the treatment or prevention of depression. Unlike the previously published meta-reviews, this review does not seek to specifically examine the evidence for particular programs but rather examines the effectiveness of unguided and guided iCBT interventions in general. The review also expands the 2011 review by including non-adult (youth) study populations, and by examining the use of iCBT for the prevention of depression (as well as for treatment). In addition, the current meta-review includes reviews which focus on participant adherence to iCBT interventions and the cost-effectiveness of iCBT for depression.

2.2 Method

The current meta-review was guided by the *Preferred Reporting Items for Systematic Reviews and Meta-Analysis 'PRISMA'* statement [21]. Database searches, data extraction and quality ratings were undertaken by a single reviewer (the candidate).

2.2.1 Inclusion and exclusion criteria for systematic review

Reviews were eligible for inclusion if they:

i) involved a systematic review of research trials of the effectiveness or cost-effectiveness of Internet-delivered CBT interventions for depression;

ii) included at least one randomised controlled trial which:

   a) examined the effectiveness or cost-effectiveness of iCBT (either guided or unguided);

   b) targeted users with depression or which aimed to prevent depression in users; and

   c) reported data relating to effectiveness, cost-effectiveness or user adherence to the intervention;

iii) were published in a peer-reviewed, English language journal; and

iv) were published after January 2000 and before August 2017.

Reviews were excluded if they were concerned with depression only as a comorbidity of a physical disorder or another mental health condition, or with depression in carers of people with a physical or mental disorder. In addition, reviews which included studies of cCBT but not studies of Internet-delivered CBT were excluded. Reviews were also excluded if they were a review of other systematic reviews.

2.2.2 Search strategy

PubMed, Cochrane Library and PsycINFO databases were searched electronically in August 2017, using the search terms listed in Table 2.1. In addition, the reference lists of reviews meeting the inclusion criteria were hand-searched for other potentially relevant reviews.

Table 2.1. Search strategy for systematic review

PubMed search terms	
1	(depressive OR depression OR dysthymia OR mood OR mental) [Title/Abstract]
2	(internet OR web OR online OR web-delivered) [Title]
3	(intervention OR treatment OR prevention OR CBT OR therapy) [Title/Abstract]
4	(systematic OR review OR meta-analysis OR metaanalysis OR meta-analytic) [Title]
5	1 AND 2 AND 3 AND 4
Cochrane Library search terms	
1	(depressive OR depression OR dysthymia OR mood OR mental) in Title, Abstract, Keywords
2	(internet OR web OR online OR web-delivered) in Record Title
3	(intervention OR treatment OR prevention OR CBT OR therapy) in Title, Abstract, Keywords
4	1 AND 2 AND 3
PsycINFO search terms	
1	(depressive OR depression OR dysthymia OR mood OR mental).mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures]
2	(internet OR web OR online OR web-delivered).m_titl.
3	(intervention OR treatment OR prevention OR CBT OR therapy).mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures]
4	(systematic OR review OR meta-analysis OR metaanalysis OR meta-analytic).m_titl.
5	1 AND 2 AND 3 AND 4

2.2.3 Data extraction

Citations and abstracts of articles resulting from the database searches were downloaded into referencing software (EndNote X8). Full-text articles were retrieved for papers which were included after abstract screening. Data were extracted from the full-text articles and recorded in an Excel spreadsheet including: focus (whether the review was focused on effectiveness, adherence and/or

cost-effectiveness); the type of trials included; the population targeted by the review; the total number of trials included in the review; how many of these trials evaluated interventions for depression which used technology; how many trials evaluated iCBT interventions for depression; whether the included trials involved guided or self-guided interventions, or both; and the results of the reviews (including overall effect sizes for quantitative reviews where calculated).

2.2.4 Quality of reviews

The quality of included systematic reviews was assessed against the *'AMSTAR' (Assessing the Methodological Quality of Systematic Reviews)* tool [22] as adapted by Kim et al. [23]. A previous systematic review of nine studies evaluating the measurement properties of the AMSTAR concluded that it was reliable and valid for use in RCT designs [24].

The adapted AMSTAR (see Appendix A: Adapted AMSTAR Checklist Items) included 11 checklist items to assess different aspects of the quality of systematic reviews, including whether the review: specified the research aims (Item 1); employed independent raters for data selection/extraction (Item 2); detailed a comprehensive search process (Item 3); clearly reported inclusion and exclusion criteria (Item 4); provided a list of included studies (Item 5); detailed the characteristics of included studies (Item 6); assessed and documented the scientific quality of included studies (Item 7); formulated conclusions based on the scientific quality of the included studies (Item 8); used appropriate methods to combine the study results (Item 9); assessed the likelihood of publication bias (Item 10); and acknowledged sources of funding and conflicts of interest (Item 11).

Articles were rated 'Yes' (Score = 1) or 'No' (Score = 0) on each of the 11 criteria according to whether the criterion was satisfied. If an item was not relevant (for example, if no quantitative analysis was undertaken), the item was marked 'Not applicable', and if an item was relevant but not reported, it was marked 'Can't answer'. A score out of 11 was computed by summing the number of items with a score of '1' (Yes). For the purposes of creating a summed score, 'Not applicable' ratings were assigned '1' and 'Can't answer' ratings were assigned '0'.

2.3 Results

2.3.1 Search results

As detailed in the study selection flow provided in Figure 2.1, the database search yielded a total of 211 citations across the three databases, of which 67 were identified as duplicates and removed. The abstracts of the remaining 144 records were then screened using the inclusion/exclusion criteria, and a further 81 records were excluded. Full-text articles were retrieved for the 62 remaining citations of which 26 failed to meet the inclusion/exclusion criteria. Half of the articles excluded at this point were reviews of trials which targeted people with health conditions other than depression, such as pain, stress, cancer, gastrointestinal conditions, coronary heart disease, insomnia, post-traumatic stress-disorder, suicidality and chronic physical illness (*n* = 13). In addition to the 36 reviews assessed as eligible from the database searches, 10 additional systematic reviews were identified through hand-searching the reference lists of eligible papers. Thus, a total of 46 systematic reviews were included in the current study.

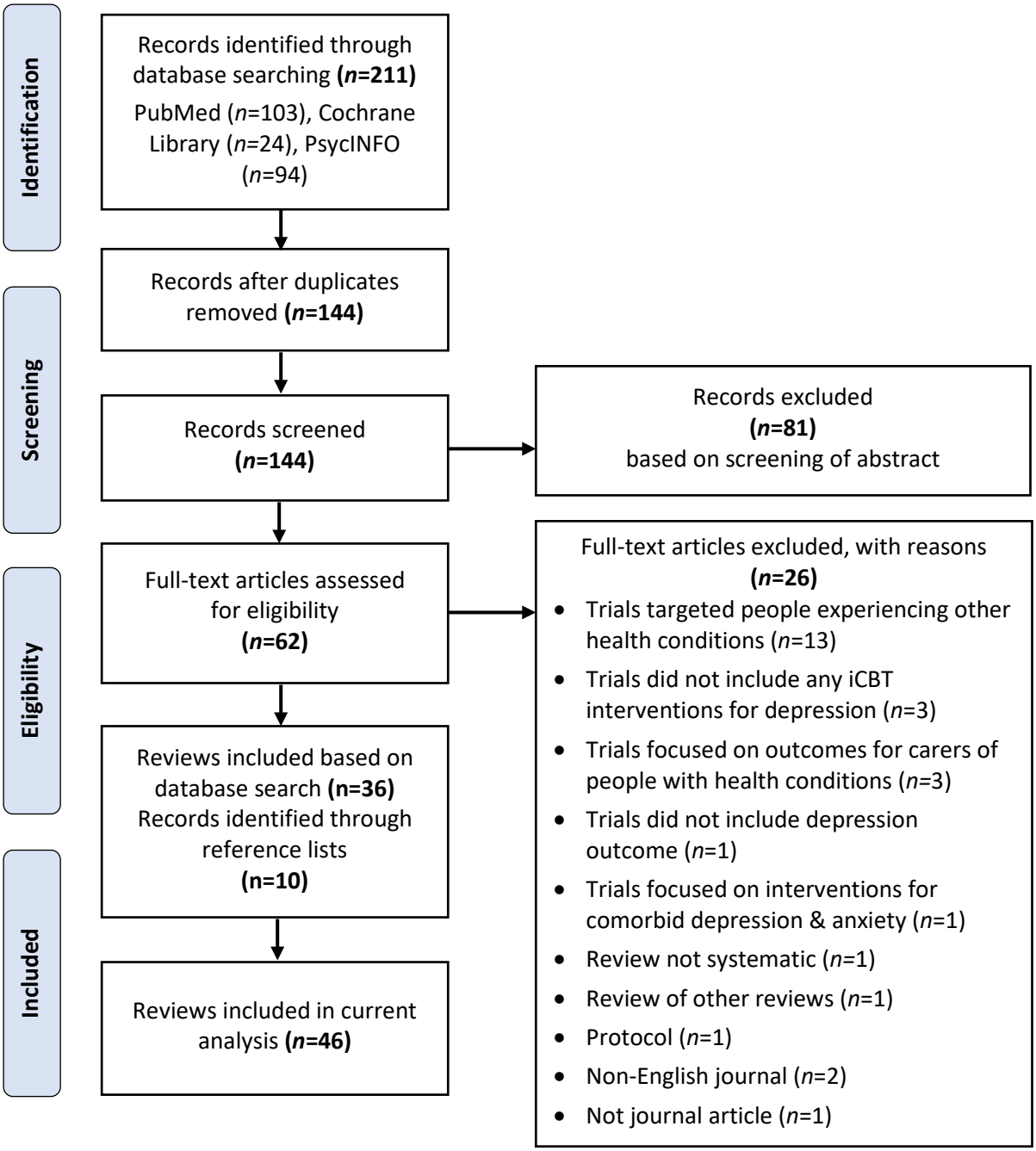

Figure 2.1. Preferred reporting items for systematic reviews and meta-analysis (PRISMA) diagram displaying procedure for selecting reviews

2.3.2 Review characteristics

Focus

Most of the 46 included reviews (*n* = 41, 89.1%) focused on the *effectiveness* of interventions in reducing or preventing depression symptoms or depressive disorder. Of these 41 systematic reviews of effectiveness, 21 were quantitative and reported a meta-analysis of the included trials. Seven reviews (15.2%) focused on participant *adherence* to interventions in the trials reviewed – four

exclusively reported on adherence and three reviews also reported on trials of effectiveness. Three reviews (6.5%) focused on *cost-effectiveness*, with two of these studies also including studies of effectiveness. Table 2.2 shows the number of reviews which focused on each of effectiveness, adherence, cost-effectiveness, or combinations of these.

Table 2.2. The focus of reviews included in the current systematic meta-review.

Focus of review:	**Number of reviews:**
Effectiveness only	36
Adherence only	4
Cost-effectiveness only	1
Effectiveness and adherence	3
Effectiveness and cost-effectiveness	2
Total	**46**

Types of trials

Nearly half of the included reviews (*n* = 21, 45.7%) were confined to studies of the *treatment* of depression, involving participants with a depression diagnosis or elevated symptoms of depression as measured by a validated depression screening test. A small number of reviews (*n* = 3, 6.5%) focused exclusively on studies of the prevention of depression (*n* = 1 indicated trials involving participants with depressive symptoms which did not meet diagnostic criteria; *n* = 1 universal trials comprising participants selected regardless of their level of depressive symptoms; and *n* = 1 trials involving a mix of indicated, selective and universal prevention trials). Three reviews (6.5%) were restricted to trials targeting perinatal populations and included both selective prevention (targeting groups at risk of depression) and treatment trials. The remaining 19 included reviews (41.3%) included both prevention and treatment trials.

Population

A large proportion of the 46 included reviews (*n* = 20, 43.5%) included studies which targeted adults only, 15.2% (*n* = 7) were focused on youth or children, and the remainder (*n = 19,* 41.3%) included studies with participants across the age spectrum.

2.3.3 Quality of reviews

The ratings for each of the 46 included reviews on the modified 11-item AMSTAR tool are displayed in Table 2.3. The average cumulative score across reviews was 9.0 (out of a possible score of 11), indicating that overall, the systematic reviews were of high quality.

All 46 reviews clarified the research purpose of the systematic review (modified Item 1), and all but one review (97.8%) described a comprehensive literature search (Item 3) and provided a list of included studies (Item 5). Most reviews detailed the characteristics of included studies (Item 6; *n* = 44, 95.7%) and clearly specified the inclusion and exclusion criteria (modified Item 4; *n* = 42, 91.3%). Just over three-quarters of the reviews (*n* = 35, 76.1%) included a conflict of interest statement and acknowledgement of funding sources (Item 11).

The use of two independent data extractors (Item 2) was reported in 30 reviews (65.2%); however, this item was assessed as 'Can't answer' for 10 studies (21.7%), for which the personnel undertaking data extraction and study selection was not described. Scientific quality of the included studies was assessed and documented (Item 7) in 71.7% of reviews (*n* = 33), with over half of these (*n* = 19) using the Cochrane Collaboration risk of bias tool [25] to assess the risk of bias in the included RCTs. Scientific quality was then subsequently considered in the analysis and reporting of results in 30 reviews (Item 8; 65.2%). All 24 of the reviews which undertook a quantitative meta-analysis used appropriate methods to combine data, in particular assessing and describing heterogeneity (Item 9). Publication bias was only assessed via graphical aids (e.g., funnel plot) or statistical tests (e.g., Egger regression test) in 18 reviews (Item 10; 39.1%).

Table 2.3. AMSTAR item ratings for systematic reviews included in the current systematic meta-review.

	1 Research purpose clear	2 Duplicate selection/ extraction	3 Wide-ranging search	4 Clear criteria	5 List of studies	6 Details of studies	7 Quality assessed	8 Quality considered	9 Suitable methods combine	10 Publication bias assessed	11 Conflict of interest stated	Sum[a]
Sztein DM, et al., 2017 [26]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	11
Lau, Y, et al., 2017 [27]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	11
Karyotaki E, et al., 2017 [28]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	11
Josephine K, et al., 2017 [29]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	10
Zhou, T, et al., 2016 [30]	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	10
Sander L, et al., 2016 [31]	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	10
Lee, EW, et al., 2016 [32]	Y	Y	Y	Y	Y	Y	Y	Y	N/A	N	Y	10
Ebert DD, et al., 2016 [33]	Y	C/A	Y	Y	Y	Y	Y	Y	Y	Y	Y	10
Brown M, et al., 2016 [34]	Y	Y	Y	Y	Y	Y	Y	N	N/A	N	Y	9
de BitenCourt MD, et al., 2016 [35]	Y	Y	Y	Y	Y	Y	N	N	N/A	N	N	7
Ashford MT, et al., 2016 [36]	Y	N	Y	Y	Y	Y	**Y**	Y	Y	N	N	8
Karyotaki E, et al., 2015 [37]	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	10

Table 2.3. AMSTAR item ratings for systematic reviews included in the current systematic meta-review (continued).

	1 Research purpose clear	2 Duplicate selection/ extraction	3 Wide-ranging search	4 Clear criteria	5 List of studies	6 Details of studies	7 Quality assessed	8 Quality considered	9 Suitable methods combine	10 Publication bias assessed	11 Conflict of interest stated	Sum[a]
Ebert DD, et al., 2015 [38]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	11
Donker, T, et al., 2015 [39]	Y	Y	Y	Y	Y	Y	Y	Y	N/A	N	Y	10
Clarke AM, et al., 2015 [40]	Y	Y	Y	Y	Y	Y	**Y**	Y	N/A	N	Y	10
Ye X, et al., 2014 [41]	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	10
van Ballegooijen, et al., 2014 [42]	Y	Y	Y	Y	Y	Y	Y	N	Y	N	Y	9
Saddichha S, et al., 2014 [43]	Y	Y	Y	Y	Y	Y	Y	Y	N/A	N	N	9
Rice S, et al., 2014 [44]	Y	Y	Y	N	Y	Y	N	N	N/A	N	Y	7
Reyes-Portillo JA, et al., 2014 [45]	Y	Y	Y	Y	Y	Y	Y	N	N/A	N	Y	9
Davies EB, et al., 2014 [46]	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	10
Baumeister H, et al., 2014 [47]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	11
Arnberg FK, et al., 2014 [48]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	11
Andersson G, et al., 2014 [49]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	11
Paul CI, et al., 2013 [50]	Y	Y	Y	Y	Y	Y	Y	Y	N/A	Y	Y	11

Table 2.3. AMSTAR item ratings for systematic reviews included in the current systematic meta-review (continued).

	1 Research purpose clear	2 Duplicate selection/ extraction	3 Wide-ranging search	4 Clear criteria	5 List of studies	6 Details of studies	7 Quality assessed	8 Quality considered	9 Suitable methods combine	10 Publication bias assessed	11 Conflict of interest stated	Sum[a]
Grist R, et al., 2013 [6]	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	N	9
Cowpertwait L, et al., 2013 [51]	Y	C/A	Y	Y	Y	Y	Y	Y	N/A	Y	N	9
Richards, D, et al., 2012 [52]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	10
Muresan, V, et al., 2012 [53]	Y	C/A	Y	N	Y	Y	N	N	Y	N	N	5
Hedman E, et al., 2012 [54]	Y	C/A	Y	Y	Y	Y	N	N	N/A	N	Y	7
Titov, N, 2011 [55]	Y	N	C/A	N	Y	Y	N	N	N/A	N	Y	5
Wade, AG, et al., 2010 [56]	Y	N	Y	Y	N	N	N	N	N/A	N	Y	5
Griffiths KM, et al., 2010 [57]	Y	C/A	Y	Y	Y	Y	Y	Y	N/A	N	Y	9
García-Lizana, et al., 2010 [58]	Y	Y	Y	N	Y	Y	Y	Y	N/A	N	Y	9
Calear AL, et al., 2010 [59]	Y	N	Y	Y	Y	Y	N	N	N/A	N	Y	7
Andrews G, et al., 2010 [8]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	11
Christensen H, et al., 2009 [60]	Y	C/A	Y	Y	Y	Y	N	N	N/A	N	Y	7
Andersson G, et al., 2009 [61]	Y	C/A	Y	Y	Y	Y	N	N	Y	Y	Y	8

Table 2.3. AMSTAR item ratings for systematic reviews included in the current systematic meta-review (continued).

	1 Research purpose clear	2 Duplicate selection/ extraction	3 Wide-ranging search	4 Clear criteria	5 List of studies	6 Details of studies	7 Quality assessed	8 Quality considered	9 Suitable methods combine	10 Publication bias assessed	11 Conflict of interest stated	Sum[a]
Amstadter AB, et al., 2009 [62]	Y	C/A	Y	Y	Y	Y	N	N	N/A	N	Y	7
Kaltenthaler E, et al., 2008 [63]	Y	Y	Y	Y	Y	Y	Y	Y	N/A	N	Y	10
Barak, A, et al., 2008 [64]	Y	Y	Y	Y	Y	Y	N	N	Y	Y	N	8
Spek V, et al., 2007 [65]	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	10
Griffiths KM, et al., 2007 [7]	Y	C/A	Y	Y	Y	Y	N	N	N/A	N	N	6
Gellatly J, et al., 2007 [66]	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	10
Kaltenthaler E, et al., 2006 [18]	Y	Y	Y	Y	Y	Y	Y	Y	N/A	N	Y	10
Griffiths KM, et al., 2006 [17]	Y	C/A	Y	Y	Y	Y	N	N	N/A	N	N	6
Total n (%)	**46 (100%)**	**30 (65.2%)**	**45 (97.8%)**	**42 (91.3%)**	**45 (97.8%)**	**44 (95.7%)**	**33 (71.7%)**	**30 (65.2%)**	**26 (100%[b])**	**18 (39.1%)**	**35 (76.1%)**	**Mean: 9**

[a] For summed score, N/A is given a score of 1 and C/A is given a score of 0.

[b] Percentage of those reviews for which this item is relevant.

Y= Yes; N= No; C/A= 'Can't Answer' This indicates the item is relevant but is not described; N/A= 'Not applicable' This indicates the item isn't relevant because a quantitative analysis wasn't conducted.

2.3.4 Effectiveness of iCBT for depression

The reviews which focused on the effectiveness of interventions for depression spanned the years 2006 to 2017 inclusive, and the details of these 41 reviews are reported in reverse chronological order in Table 2.4.

In the earliest reviews of Internet interventions for common mental disorders, Kaltenthalter and colleagues identified two RCTs of iCBT for depression in adults (one of which reported positive results) [18], and Griffiths and Christensen reported on three RCTs with non-active controls for any age group (one with positive results for depression symptoms) [17]. In subsequent updates to these reviews in 2008 [63] and 2007 [7], the number of RCTs had increased to four (three with positive results) and four (two with positive results) respectively. In an additional 2010 update to their review of interventions for all age groups, Griffiths et al. [57] noted a linear growth in the number of RCTs of Internet interventions for depression and anxiety and reported on eight RCTs of interventions for depression. Six of these RCTs reported a significant effect for the iCBT group compared to a control, and two RCTs reported no effect compared to a psycho-educational control.

This growth in RCTs has continued, and in 2014, Saddichha et al. identified 29 RCTs of online interventions targeting depression, of which 25 involved CBT techniques (all ages) [43]. All studies except two reported a significant reduction in depression symptoms in the iCBT group. The effect sizes for trials of guided interventions ranged from 0.6 to 1.9, with trials of self-guided iCBT yielding lower effect sizes of 0.3 to 0.7.

Table 2.4. Characteristics of systematic reviews examining effectiveness included in the current systematic meta-review.

Review	Included studies	Population [Study types[a]]	Analysis[b]	Studies in review (N)	Dep RCTs[c] (n)	Dep iCBT RCTs[d] (n)	Support types[e]	Pooled depression outcome effect sizes[f] / conclusions of qualitative reviews
Sztein DM, et al., 2017 [26]	RCTs of iCBT for adults experiencing mild to moderate depression, compared to wait list control.	Adults with depression symptoms [Treatment]	QN	14	14	14	G, SG	Overall effect size: SMD=0.74, CI: 0.62-0.86. Guided (12 studies) SMD=0.73, CI: 0.58-0.87. Self-guided (4 studies) - SMD=0.79 CI: 0.55-1.03. No significant difference between groups. 3-6-month follow-up SMD=0.83, CI: 0.69-0.98
Lau Y, et al., 2017 [27]	RCTs of guided iCBT for adult women less than 2 years postpartum, for depression, pregnancy loss and post-traumatic stress.	Adult women <= 2 years post-partum [Selective]	QN	8	5	5	G	Depression iCBT (5 studies): Overall effect size: SMD=0.65, CI: 0.48-0.82
Karyotaki E, et al., 2017 [28]	RCTs of self-guided iCBT for depression.	Adults with depression (diagnosis or elevated symptoms) [Treatment]	QN	16	16	16	SG	Traditional meta-analysis (16 RCTs): Overall effect size: SMD=0.33, CI: 0.19-0.46 Individual participant meta-analysis (data from 13 RCTs): Significant effect of treatment over control: $\beta = -0.21$; $p<.001$ Overall effect size: SMD=0.27, CI: 0.17-0.37
Josephine K, et al., 2017 [29]	RCTs of Internet and mobile interventions for adults with a diagnosed depressive disorder.	Adults with depression diagnosis [Treatment]	QN	19	19	12	G, SG	10 RCTs of interventions vs wait-list (includes non-iCBT) had depression outcomes: Overall effect size: SMD=0.90, CI: 0.73-1.07

Table 2.4. Characteristics of systematic reviews examining effectiveness included in the current systematic meta-review (continued).

Review	Included studies	Population [Study types[a]]	Analysis[b]	Studies in review (N)	Dep RCTs[c] (n)	Dep iCBT RCTs[d] (n)	Support types[e]	Pooled depression outcome effect sizes[f] / conclusions of qualitative reviews
Zhou T, et al., 2016 [30]	RCTs of iCBT for subthreshold depression.	All ages with subthreshold depression symptoms (no diagnosis) [Indicated]	QN	8	8	8	G, SG	Post-intervention: Overall effect size: SMD=0.46, CI: 0.22-0.70 3-month follow-up (3 RCTs): SMD=0.12, CI: 0.01-0.22 6-month follow-up (2 RCTs) - no significant difference. 12-month follow-up (2 RCTs): SMD=0.28, CI: 0.12-0.44
Sander L, et al., 2016 [31]	RCTs of online psychological interventions in preventing mental disorders	Adults without mental disorder diagnosis [Universal]	QN	17	6	3	G, SG	8 RCTs reported depression severity outcomes (includes non-iCBT interventions): Overall effect size: SMD=0.35, CI: 0.12-0.57 At medium term follow-up: Overall effect size: SMD=0.22, CI: 0.07-0.37
Lee, EW, et al., 2016 [32]	Studies of online interventions with maternal mental health as an outcome measure.	Women, from pregnancy to 1-year post-partum [Selective]	QL	3	2	2	G, SG	Significant improvement in depression symptoms for intervention groups compared to control groups.

Table 2.4. Characteristics of systematic reviews examining effectiveness included in the current systematic meta-review (continued).

Review	Included studies	Population [Study types[a]]	Analysis[b]	Studies in review (N)	Dep RCTs[c] (n)	Dep iCBT RCTs[d] (n)	Support types[e]	Pooled depression outcome effect sizes[f] / conclusions of qualitative reviews
Ebert DD, et al., 2016 [33]	Individual participant meta-analysis based on RCTs of guided Internet interventions for adults with depression.	Adults with depression (diagnosis or elevated symptoms) [Treatment]	QN	18	18	12	G	Risk of deterioration from baseline to post-intervention significantly lower in intervention compared to control groups (includes non-iCBT interventions): RR=0.47, CI: 0.29-0.75. Difference in risk of deterioration between groups not significant at follow-up.
de BitenCourt MD, et al., 2016 [35]	Studies of psychotherapy including virtual contact with a therapist.	Adults [Mixed]	QL	35	6	3	G	2 RCTs showed significant reduction of symptoms in iCBT group, 1 RCT showed no significant difference between intervention and control group.
Ashford MT, et al., 2016 [36]	Studies of online interventions for depression, stress and complicated grief for women in perinatal period.	Women, from pregnancy to 1-year post-partum [Selective]	QN	11	2	1	G, SG	Meta-analysis of 8 RCTs (includes depression as secondary outcome, non-iCBT interventions): Significant effect for intervention in 6 studies. 2 studies had non-significant difference (although interventions targeted at stress). Overall effect size not significant.

Table 2.4. Characteristics of systematic reviews examining effectiveness included in the current systematic meta-review (continued).

Review	Included studies	Population [Study types[a]]	Analysis[b]	Studies in review (N)	Dep RCTs[c] (n)	Dep iCBT RCTs[d] (n)	Support types[e]	Pooled depression outcome effect sizes[f] / conclusions of qualitative reviews
Ebert DD, et al., 2015 [38]	RCTs of Internet and computerised CBT for depression and anxiety in youth.	Youth <=25 years with depression/ anxiety (diagnosis or elevated symptoms) [Treatment]	QN	13	4	4	G, SG	For 4 RCTs on depression iCBT: Overall effect size SMD=0.76, CI: 0.41-1.12.
Clarke AM, et al., 2015 [40]	Studies of interventions to promote positive mental health or prevent mental health problems in youth.	Youth aged 12-25 years [Mixed prevention]	QL	28: 8 promotion 20 prevention	6	6	G, SG	5 RCTs: Significant reduction in depression in iCBT compared to control group, at post-intervention and follow-up. 1 RCT: Significant reduction in depression in iCBT group in male participants only, at post-intervention and 6-month follow-up
Ye X, et al., 2014 [41]	Studies of Internet interventions targeting anxiety and/or depression symptoms, for youth.	Youth aged 7-25 years [Mixed]	QN	7	3	2	G, SG	For all studies: Overall intervention effect on depression symptoms not significant, however this included non-depression interventions and non-iCBT interventions. Overall effect size SMD=0.16, CI: -0.12-0.44

Table 2.4. Characteristics of systematic reviews examining effectiveness included in the current systematic meta-review (continued).

Review	Included studies	Population [Study types[a]]	Analysis[b]	Studies in review (N)	Dep RCTs[c] (n)	Dep iCBT RCTs[d] (n)	Support types[e]	Pooled depression outcome effect sizes[f] / conclusions of qualitative reviews
Saddichha S, et al., 2014 [43]	Studies of online self-help interventions targeting depression or anxiety.	No restriction [Mixed]	QL	33	29	25	G, SG	Guided CBT: Effect sizes from 0.6 to 1.9. Self-guided CBT: Effect sizes from 0.3 to 0.7. All RCTs of iCBT for depression: positive results, except 1 RCT where no significant difference for depression (unguided iCBT) and for 1 RCT where iCBT was less effective than a problem-solving intervention.
Rice S, et al., 2014 [44]	RCTs of interventions for treatment or prevention of depression.	Youth aged 12-25 years [Mixed]	QL	15	15	15	G, SG	9 depression prevention RCTs - effective for university students, adolescents (selective) and secondary students (significant effect for males only). 5 depression intervention RCTs – effective (1 RCT showed non-inferiority compared to TAU) 1 combined intervention: no significant effect.
Reyes-Portillo JA, et al., 2014 [45]	Studies of interventions for treatment or prevention of depression.	Youth aged <=25 years [Mixed]	QL	17	13	10	G, SG	4 RCTs of depression iCBT showed significant depression reductions at post-intervention, with effect sizes from 0.15 to 0.84. 6 RCTs of depression iCBT showed no significant effect on depression at post-test for iCBT group compared to control.

Table 2.4. Characteristics of systematic reviews examining effectiveness included in the current systematic meta-review (continued).

Review	Included studies	Population [Study types[a]]	Analysis[b]	Studies in review (N)	Dep RCTs[c] (n)	Dep iCBT RCTs[d] (n)	Support types[e]	Pooled depression outcome effect sizes[f] / conclusions of qualitative reviews
Davies EB, et al., 2014 [46]	RCTs of Internet interventions for depression, anxiety, distress and stress in university students.	Higher education students [Mixed]	QN	17	9	8	G, SG	For 9 RCTs of depression interventions (includes 1 non-iCBT): Overall effect size SMD=0.43, CI: 0.22-0.63.
Baumeister H, et al., 2014 [47]	RCTs of Internet interventions, comparing guided vs unguided, guidance intensity or coach type.	Adults with mental disorder [Treatment]	QN	14	3	3	G vs SG	3 RCTs of guided vs unguided iCBT for depression: Overall effect size: SMD= - 0.15, CI: -0.46-0.16 (non-significant but favouring guided).
Arnberg FK, et al., 2014 [48]	RCTs of Internet interventions for anxiety and mood disorders. Trials excluded if high risk of bias.	All ages with a diagnosis of anxiety or mood disorder [Treatment]	QN	40	9	5	G, SG	iCBT was guided in the 5 depression iCBT RCTs. Guided iCBT compared to wait-list (5 RCTs): Overall effect size SMD=0.83, CI: 0.59-1.07.
Andersson G, et al., 2014 [49]	Studies of guided iCBT compared to face-to-face CBT for mental disorders	No restriction [Treatment]	QN	13	2	2	G	For 2 depression RCTs iCBT vs F2F: Overall effect size not significant SMD=0.05, CI: -0.19-0.30.
Paul CI, et al., 2013 [50]	Studies of web-based interventions for chronic mental and physical health conditions.	No restriction [Treatment]	QL	36	12	9	G, SG	For 9 RCTs of iCBT for depression: 7 showed significant improvement in depression in intervention group compared to control. 2 RCTs showed no significant intervention effect.

Table 2.4. Characteristics of systematic reviews examining effectiveness included in the current systematic meta-review (continued).

Review	Included studies	Population [Study types[a]]	Analysis[b]	Studies in review (N)	Dep RCTs[c] (n)	Dep iCBT RCTs[d] (n)	Support types[e]	Pooled depression outcome effect sizes[f] / conclusions of qualitative reviews
Grist R, et al., 2013 [6]	RCTs of CCBT for common mental disorders, compared to other therapies or wait-list controls.	Adults >16 years with a common mental disorder [Treatment]	QN	49	14	14	G, SG	For all 49 RCTs (incl. non-depression outcomes): Overall effect size SMD=0.77, 0.59-0.95. No significant difference in effect sizes for type of problem or for amount of therapeutic support.
Cowpertwait L, et al., 2013 [51]	RCTs of web-based interventions for depression	Adults [Treatment]	QN	18	18	unclear	G, SG	Overall effect size SMD=0.43, CI: 0.29-0.57. RCTs of guided interventions had sig larger effect sizes SMD=0.48 than self-guided SMD=0.32.

Table 2.4. Characteristics of systematic reviews examining effectiveness included in the current systematic meta-review (continued).

Review	Included studies	Population [Study types[a]]	Analysis[b]	Studies in review (N)	Dep RCTs[c] (n)	Dep iCBT RCTs[d] (n)	Support types[e]	Pooled depression outcome effect sizes[f] / conclusions of qualitative reviews
Richards D, et al., 2012 [52]	Studies of computer-based interventions for adults with depression	Adults with depression (diagnosis or elevated symptoms) [Treatment]	QN	45 (systematic review) 23 (meta-analysis - 19 trials)	19	17	G, SG	For 19 RCTs for depression (includes 2 non-iCBT) Post-intervention: Overall effect size SMD=0.56, CI: 0.41-0.71. Follow-up (14 RCTs): Overall effect size SMD=0.20, CI: 0.09-0.31. Guided (therapist) (7 RCTs): Overall effect size SMD=0.78, CI: 0.64-0.92. Guided (administrative) (5 RCTs): Overall effect size SMD=0.58, CI: 0.28-0.88. Self-guided (9 RCTs): Overall effect size SMD=0.36, CI: 0.10-0.61. NSD between SG and admin support. Significant difference between SG and therapist support.
Muresan V, et al., 2012 [53]	RCTs of iCBT where cognition was measured.	Clinical or sub-clinical depression or anxiety symptoms [Treatment, Indicated]	QN	11	3	2	G, SG	For 9 RCTs where depression measured (includes depression as secondary outcome): Overall effect size (weighted mean to account for variation in sample size) SMD=0.70 CI: 0.53-0.88.

Table 2.4. Characteristics of systematic reviews examining effectiveness included in the current systematic meta-review (continued).

Review	Included studies	Population [Study types[a]]	Analysis[b]	Studies in review (N)	Dep RCTs[c] (n)	Dep iCBT RCTs[d] (n)	Support types[e]	Pooled depression outcome effect sizes[f] / conclusions of qualitative reviews
Hedman E, et al., 2012 [54]	RCTs of iCBT for a clinical problem	Adults [Treatment, Indicated]	QL	103	20	20	G, SG	Within-group effect sizes for depression symptoms in iCBT trial arms ranged from 0.38 to 2.27 with a mean of 0.94, CI: 0.77-1.11.
Titov N, 2011 [55]	RCTs of Internet interventions for depression	Adults with depression (diagnosis or elevated symptoms) [Treatment]	QL	13	13	12	G, SG	4 'low intensity' guided iCBT RCTs: effect sizes 0.50 to 1.09. 3 'high intensity' guided iCBT RCTs: effect sizes 0.70 to 0.96. 5 self-guided iCBT RCTs: effect sizes 0.02 to 0.55.
Wade AG, 2010 [56]	Studies of use of Internet for screening, education and treatments	Any age with anxiety or depression [Treatment]	QL	75	unclear	5	G, SG	4 iCBT RCTs for depression: significant positive effect for iCBT compared to control group. 1 iCBT RCT for depression: no effect between iCBT and control.
Griffiths KM, et al., 2010 [57]	RCTs of Internet interventions for depression or anxiety, with non-active control	No restriction [Mixed]	QL	26	8	8	G, SG	6 iCBT RCTs for depression: significant positive effect for iCBT compared to control (including 1 attention control). 2 RCTs of iCBT for depression: no effect for iCBT compared to psycho-educational control.

Table 2.4. Characteristics of systematic reviews examining effectiveness included in the current systematic meta-review (continued).

Review	Included studies	Population [Study types[a]]	Analysis[b]	Studies in review (N)	Dep RCTs[c] (n)	Dep iCBT RCTs[d] (n)	Support types[e]	Pooled depression outcome effect sizes[f] / conclusions of qualitative reviews
García-Lizana F, et al., 2010 [58]	RCTs of interventions using technology for management of depression	No restriction [Treatment, Indicated]	QL	10	10	4	G, SG	3 iCBT RCTs for depression: significant positive results for iCBT compared to control. 1 iCBT RCT for depression: no effect.
Calear AL, et al., 2010 [59]	Studies of Internet-based prevention and treatment of anxiety and depression	Children and youth aged 5-19 years [Mixed]	QL	8	6	1	G, SG	1 iCBT RCT for depression - effective for males only (d=0.43 post-intervention; d=0.27 follow-up)
Andrews G, et al., 2010 [8]	RCTs of iCBT and CCBT for anxiety and depressive disorders	Adults with diagnostic criteria for disorders [Treatment]	QN	22	6	5	G, SG	For all depression RCTs – 5 iCBT (all guided), 1 CCBT (self-guided): Overall effect size SMD=0.78, CI: 0.59-0.96.
Andersson G, et al., 2009 [61]	RCTs of Internet or other computerised psychological treatments for depression	Adults with depression (diagnosis or elevated symptoms) [Treatment]	QN	12	12	9	G, SG	12 out of 15 comparisons in the 12 RCTs were for CBT treatments (iCBT or CCBT): Overall effect size SMD=0.42, CI: 0.26-0.59. For all RCTs: Difference between effects sizes for guided (SMD=0.62) vs self-guided (SMD=0.25) was significant.
Amstadter AB, et al., 2009 [62]	RCTs of Internet interventions for traumatic stress-related mental health problems (including depression)	No restriction [Treatment, Indicated]	QL	unclear	10	9	G, SG	8 of 9 iCBT RCTs of depression reported significant positive results for iCBT compared to control.

Table 2.4. Characteristics of systematic reviews examining effectiveness included in the current systematic meta-review (continued).

Review	Included studies	Population [Study types[a]]	Analysis[b]	Studies in review (N)	Dep RCTs[c] (n)	Dep iCBT RCTs[d] (n)	Support types[e]	Pooled depression outcome effect sizes[f] / conclusions of qualitative reviews
Kaltenthaler E, et al., 2008 [63]	RCTs of CCBT for depression (includes computer-based interventions)	Adults with mild to moderate depression [Treatment]	QL	4	4	4	G, SG	3 of 4 iCBT RCTs reported significant positive results for iCBT compared to control.
Barak A, et al., 2008 [64]	Studies of Internet delivered therapy for any condition	No restriction [Treatment, Indicated]	QN	92 studies described in 64 articles	6	6	G, SG	For depression symptom effects measured in any trial (n=16): Overall effect size: SMD=0.32
Spek V, et al., 2007 [65]	RCTs of iCBT for depression or anxiety, with non-active control	Adults [Mixed]	QN	12	5	5	G, SG	For RCTs of depression iCBT (1 guided, 4 self-guided): Overall effect size SMD=0.27, CI: 0.15-0.40.
Griffiths KM, et al., 2007 [7]	RCTs of Internet interventions for mental disorders, with non-active control	No restriction [Mixed]	QL	23	4	4	G, SG	2 of 4 iCBT RCTs reported significant positive results for iCBT compared to control.
Gellatly J, et al., 2007 [66]	RCTs of self-help for depression versus controls (includes non-technical interventions)	No restriction [Mixed]	QN	34 (39 comparisons)	10 compar-isons	unclear	G, SG	Overall (all self-help interventions): SMD=0.43, CI: 0.30-0.57. Multivariate regression: Technical vs non-technical - NSD. G significantly more effective than SG. CBT significantly more effective than non-CBT.

Table 2.4. Characteristics of systematic reviews examining effectiveness included in the current systematic meta-review (continued).

Review	Included studies	Population [Study types[a]]	Analysis[b]	Studies in review (N)	Dep RCTs[c] (n)	Dep iCBT RCTs[d] (n)	Support types[e]	Pooled depression outcome effect sizes[f] / conclusions of qualitative reviews
Kaltenthaler E, et al., 2006 [18]	Studies of CCBT for treatment of anxiety, depression, phobias, panic and OCD (includes computer-based)	Adults with depression [Treatment]	QL	20	10	2	G, SG	1 of 2 iCBT RCTs reported significant positive results for iCBT compared to control.
Griffiths KM, et al., 2006 [17]	RCTs of Internet interventions for mental disorders, with non-active control	No restriction [Mixed]	QL	15	3	3	G, SG	1 of 3 iCBT RCTs (an indicated prevention study) reported positive results for iCBT compared to control.

[a] Type of studies included, based on intervention or population criteria: Treatment, Indicated (prevention), Selective (prevention), Universal (prevention), Mixed (treatment and prevention), Mixed prevention (indicated, selective and universal prevention).

[b] QN= Quantitative meta-analysis, QL= Qualitative review.

[c] Number of RCTs examining a depression intervention which uses technology.

[d] Number of RCTs examining a depression iCBT intervention.

[e] G= Guided support (either therapist or administrative support), SG=Self-guided.

[f] Between-group effect sizes for depression outcomes unless otherwise specified.

iCBT=Internet-delivered cognitive behaviour therapy; RCT=Randomised Controlled Trial; SMD= Standardised mean difference. All SMDs are significant unless otherwise specified; CI= 95% Confidence Interval; RR=Relative Risk; F2F=Face-to-face; CCBT= Computerised cognitive behaviour therapy; NSD=No significant difference.

The following sections consider the results of the reviews of (a) treatment; (b) prevention; and (c) other intervention studies for depression, separately presenting the findings for reviews targeting different population sub-groups.

(a) Reviews of treatment studies (for depression diagnosis or elevated symptoms)

Of the depression outcome effectiveness reviews, 19 focused on the treatment of participants with a diagnosis of depression or with elevated symptoms of depression as measured by a validated depression screening test.

Adults only: Of the reviews of treatment trials, 14 were restricted to adults with depression [6, 8, 18, 26, 28, 29, 33, 47, 51, 52, 54, 55, 61, 63]. Seven were specifically focused on CBT interventions [6, 8, 18, 26, 28, 47, 54, 63], whereas the remainder included trials of non-CBT interventions (involving other therapeutic techniques such as mindfulness and physical activity) as well as trials of iCBT.

The majority of these reviews reported positive intervention effects for iCBT compared to control groups. Eight quantitative meta-analyses reported effect sizes calculated from pooled data ranging from 0.33 to 0.90 [6, 8, 26, 28, 29, 51, 52, 61]. In the most recent of these quantitative reviews, which was assessed as having a quality of score of 11 (maximum possible), Sztein et al. identified 14 RCTs where an iCBT intervention was compared to a wait-list control [26]. The overall effect size for depression outcomes measured at post-intervention reported by these RCTs was moderate to large (*SMD* = 0.74, 95% CI: 0.62, 0.86), and this effect was sustained at longer-term follow-up (3 to 6 months after the intervention) with an overall effect size of 0.83 (95% CI: 0.69, 0.98).

Many reviews separated results for trials of guided and unguided interventions, with overall effect sizes calculated for guided and unguided programs for depression in four quantitative reviews [26, 51, 52, 61]. In the earliest of these reviews, published in 2009 and including 12 RCTs, Andersson and Cuijpers reported a significant difference between guided interventions (*SMD* = 0.61, 95% CI: 0.45, 0.77) and unguided interventions (*SMD* = 0.25, CI: 0.14, 0.35). However, this analysis included trials of computerised (not Internet) CBT, as well as trials of Internet interventions which were not CBT-based. Similarly, in a review published in 2013, Cowpertwait and Clarke who updated and extended the Andersson review, undertook a meta-analysis of 18 RCTs of web-based interventions (but not limited to CBT) and found a significant difference between guided (*SMD* = 0.48, 95% CI: 0.39, 0.57) and unguided (*SMD* = 0.32, 95% CI: 0.23, 0.41) interventions [51].

The comparative effectiveness of guided and unguided treatment interventions was also examined in detail by Richards and Richardson in a 2012 meta-analysis of 19 RCTs although the analysis included

non-CBT as well as CBT interventions, and non-Internet- as well as Internet-based CBT interventions [52]. The review included seven RCTs involving therapist support, which had an overall effect size of 0.78 (95% CI: 0.64, 0.92). Five RCTs involved 'administrative' support (provided by a non-clinical staff member) and had an overall effect size of 0.58 (95% CI: 0.28, 0.88), and nine RCTs involved unguided interventions with no support which had an overall effect size of 0.36 (95% CI: 0.10, 0.61). The depression outcomes for trials of therapist supported interventions were significantly better than those for unguided interventions; however, the difference between outcomes for administrative supported and unguided interventions was not significant.

In contrast to these results, the most recent review, published in 2017 and which focused specifically on iCBT, found no significant difference between the results for guided and unguided interventions [26]. The review included 12 studies of guided iCBT (*SMD* = 0.73, 95% CI: 0.58, 0.87) and four studies of unguided iCBT (*SMD* = 0.79, 95% CI: 0.55, 1.03). This finding is consistent with that of an earlier review published in 2014 which focused specifically on treatment studies which directly examined the comparative effectiveness of guided versus unguided Internet interventions, including three studies of iCBT for depression in adults [47]. In this meta-analysis, Baumeister and colleagues found that outcome differences in depression symptom severity between guided and non-guided iCBT were not statistically significant, although the direction of the effect favoured guided interventions (*SMD* = -0.15, 95% CI: -0.46, 0.16).

The effectiveness of *unguided* iCBT for depression in adults was the exclusive focus of a review published in 2017 by Karyotaki and colleagues [28]. The review included studies where the iCBT intervention was not guided by a therapist and was compared to either a usual care, wait-list or attention control, resulting in 16 eligible studies. The overall effect size for these studies was small to moderate (*SMD* = 0.33, 95% CI: 0.19, 0.46). An individual participant meta-analysis was then performed using data from 13 of the 16 trials for which data was available. This analysis showed a significant treatment effect of unguided iCBT over control conditions ($\beta = -0.21$; $p < .001$) and a small effect size (*SMD* = 0.27, 95% CI: 0.17, 0.37).

Most reviews of effectiveness focused on improvement in outcomes. However, one 2016 review by Ebert et al. undertook an individual data meta-analysis to investigate the effect of guided Internet interventions on the deterioration in the mood of participants, designated as an increase in depressive symptoms of more than 7.68 and 7.63 points in the CES-D and BDI respectively [33]. They concluded that the intervention condition was associated with a significant reduction in risk of deterioration relative to the control condition.

No age restriction (adults and youth): Four reviews focused on studies of users with either a depression diagnosis or elevated symptoms of depression but did not restrict studies to those involving adults [48, 49, 50, 56]. Two of these reviews were qualitative; both reported that a majority of the included RCTs for depression had a significant treatment effect [50, 56]. Arnberg and colleagues reached a similar conclusion in a 2014 quantitative review of RCTs of Internet interventions of participants with an anxiety or depression diagnosis. Of the nine depression intervention studies identified by the study [48], five RCTs involved iCBT interventions, all of which were guided iCBT compared to a wait-list control. The overall effect size for these guided iCBT trials was 0.83 (95% CI: 0.59, 1.07).

One review investigated if guided iCBT is as effective as face-to-face CBT for the treatment of mental disorders [49]. The review, published in 2014 by Andersson et al., identified only two relevant depression studies and found no significant difference between the iCBT and face-to-face conditions (*SMD* = 0.05, 95% CI: -0.19, 0.30).

Youth (25 years and younger): Only one review, published in 2016, focused solely on the effectiveness of iCBT interventions for treating young people with depression. It involved a quantitative meta-analysis of RCTs of iCBT for depression and anxiety in youth with a depression or anxiety diagnosis or elevated symptoms. It identified four relevant RCTs and reported an overall effect size of 0.76 (95% CI: 0.41, 1.12) [38].

(b) Reviews of depression prevention studies

Studies evaluating the effectiveness of interventions for *prevention* can be classified as: **indicated** (including participants who have elevated depressive symptoms but who do not meet criteria for diagnosis); **selective** (targeting participants who belong to groups that are at risk of developing the disorder); or **universal** (comprising all participants regardless of depressive symptomatology) [67]. In the current meta-review, three reviews were identified which focused exclusively on studies of the prevention of depression using iCBT [30, 31, 40].

Adults only: Two reviews evaluated the effectiveness of CBT Internet prevention interventions for adults. One focused specifically on indicated intervention trials [30] while the second examined the effectiveness of universal interventions [31]. The current study did not identify any review that focused exclusively on the effect of iCBT selective prevention interventions for adult groups at risk of depression.

In their 2016 review, Zhou et al. included studies that specifically targeted participants with sub-threshold depression symptoms (but not diagnosis) [30]. The overall effect size at post-intervention for the nine included RCTs (of both guided and unguided iCBT) was moderate (*SMD* = 0.46, 95% CI: 0.22, 0.70).

In the only review of the effectiveness of universal preventive interventions, Sander and colleagues' 2016 paper identified three relevant RCTs focused on Internet interventions for the prevention of depression in adults [31]. Their meta-analysis showed a small but significant reduction in depression symptoms for the intervention groups; however, this analysis included an RCT of a mindfulness-based Internet program (*SMD* = 0.35, 95% CI: 0.12, 0.57).

Youth (25 years and younger): One review of prevention studies (of all types) of iCBT for youth aged 12 to 25 years was identified. In this review, Clarke et al. found six RCTs of iCBT for depression prevention, with five RCTs reporting a significant positive outcome for the intervention group at post-intervention and follow-up, and the remaining RCT finding a significant improvement in depression outcomes only for male participants [40].

(c) Other reviews of intervention studies (no symptom restriction)

The remaining 19 reviews of the effectiveness of iCBT did not impose any restriction on participant symptom levels in the studies reviewed.

Adults only: Five reviews examined the use of iCBT for depression in adults, without any symptom restriction [27, 32, 35, 36, 65].

The earliest of these was published in 2007 and identified five RCTs of iCBT for depression (one guided and four self-guided interventions) [65]. The overall effect size for these trials was small (*SMD* = 0.27, 95% CI: 0.15, 0.40). A more recent 2016 qualitative review was restricted to trials of Internet interventions which included some kind of therapist guidance. Of the three RCTs identified, two showed a significant reduction in depression symptoms and one showed no difference in symptom change between the intervention and control groups [35].

The three remaining reviews in this sub-category of intervention studies were concerned with interventions for women in the perinatal period (from pregnancy to one- or two-years post-partum) [27, 32, 36]. These reviews included treatment studies as well as trials for the prevention or reduction of sub-clinical depression symptoms. Lau et al. reviewed guided iCBT interventions for depression, pregnancy and post-traumatic stress, and identified five iCBT interventions for depression [27]. The overall effect size for the depression intervention groups was moderate to large

(*SMD* = 0.65, 95% CI: 0.48, 0.82). Similarly, Lee and colleagues reported significant positive improvements in the iCBT group for two identified RCTs of interventions for women up to one-year post-partum [32]. Finally, Ashford and colleagues reviewed studies of online interventions targeting depression, stress and complicated grief among women in the perinatal period and identified 11 studies [36]. They undertook a meta-analysis of the eight RCTs where depression outcomes were reported and did not find a significant overall treatment effect. However, the analysis included non-iCBT interventions and RCTs where depression was a secondary outcome.

No age restriction (adults and youth): Nine reviews included studies regardless of participant age and depression status [7, 17, 43, 53, 57, 58, 62, 64, 66] of which most (*n* = 7 reviews) were published in 2010 or earlier [7, 17, 57, 58, 62, 64, 66]. The earliest of these were qualitative reviews and included limited numbers of RCTs of iCBT, with half or less reporting positive results [7, 17, 66]; however, the number of positive RCTs markedly increased in qualitative reviews from 2008 onwards [57, 58, 62]. The most recent review published by Saddichha et al. included 25 RCTs of online CBT self-help interventions which reported depression outcomes, of which all except two reported positive outcomes [43].

Two reviews undertook a quantitative meta-analysis [43, 53, 64]. The first of these, published in 2008, included studies (not only RCTs) of any Internet delivered therapy programs [64]. For the 16 studies which measured depression symptoms the overall effect size was small (*SMD* = 0.32) [64]. A later 2012 review included both indicated prevention and treatment RCTs of iCBT for depression or anxiety in populations of any age [53]. Across all included RCTs which reported a depression outcome (including as a secondary outcome, *n* = 9), a medium to large overall effect size was reported (*SMD* = 0.70, 95% CI: 0.53, 0.88).

Youth (25 years and younger): Five reviews targeted youth aged 25 years or younger and included both treatment and prevention trials [41, 44, 45, 46, 59]. The earliest was published in 2010 and included trials for anxiety and depression, in children and adolescents aged 5 to 19 years [59]. Only one RCT for depression was included, which had a positive effect for male participants only.

The other four reviews targeting young people were all published in 2014 [41, 44, 45, 46] but yielded somewhat conflicting findings. The most inclusive was a qualitative review undertaken by Rice and colleagues which identified 15 RCTs of online interventions for the prevention or treatment of depression in participants aged 12 to 25 years [44]. The review included nine prevention trials, with positive results reported for all of these including for the use of iCBT with university students, male secondary students and adolescents at risk of depressive disorder. Four treatment intervention trials

also showed significant reductions in symptoms for the intervention group compared to a control condition, and one RCT employing a non-inferiority design demonstrated equivalence between iCBT and treatment as usual. Only one study, a combined prevention and treatment trial, reported no significant intervention effect. Similarly, another quantitative review, which focused on students in higher education, yielded a positive outcome overall for this age group. The review, which identified nine RCTs of depression interventions compared to inactive control conditions [46], reported a pooled effect size for these RCTs of 0.43 (95% CI: 0.22, 0.66), although the review included one RCT of a non-iCBT intervention.

In a less encouraging review, Reyes-Portillo et al. included studies of interventions for the treatment or prevention of depression, anxiety or suicide, in young people aged 25 years or younger [45]. The authors identified 10 RCTs of iCBT for depression but reported significant intervention effects for only four of these trials. Further, Ye and colleagues found little evidence of the effectiveness of interventions relative to control in their quantitative analysis of studies of Internet programs for anxiety and/or depression in young people [41]. The latter review identified seven relevant studies each of which was rated to have high or moderate quality and found no significant difference between the Internet intervention conditions and control condition, although the review also found no significant difference between Internet and face-to-face conditions. However, Ye et al.'s analysis was not restricted to trials of iCBT interventions and many of the programs were not specifically targeted at depression.

2.3.5 Adherence to iCBT for depression

Participant adherence to iCBT interventions refers to the extent to which a participant engages with an iCBT program. Non-adherence is equivalent to treatment drop-out in face-to-face settings [68]. In the Internet literature, adherence to online interventions is typically measured by analysing the number of logins made a by the user, the number of modules accessed or how much time a user spends using the intervention [60]. It should be noted that adherence, as defined and used in this thesis, does not refer to drop-out from a research protocol (e.g., where a participant fails to complete a post-intervention or follow-up outcome measure).

Seven reviews examined user adherence to iCBT interventions and are detailed in Table 2.5. Five of these reviews were restricted to adults with either a depression diagnosis or elevated depression symptoms (treatment trials) [37, 42, 47, 51, 52] and two reviews imposed no restriction on the depression status or age of the population under study [34, 60].

Table 2.5. Characteristics of systematic reviews examining adherence included in the current systematic review.

Review	Included studies	Population [Study types[a]]	Analysis[b]	Studies in review (N)	Dep RCTs[c] (n)	Dep iCBT RCTs[d] (n)	Support types[e]	Adherence rates / conclusions of qualitative reviews
Brown M, et al., 2016 [34]	RCTs of Internet interventions for CMDs which include gamification features.	No restriction [Mixed]	QL	61	30	Unclear	G, SG	For 45 RCTs (includes non-depression): Completion rate = 54.0% (SD: 24.6%)
Karyotaki E, et al., 2015 [37]	RCTs of self-guided web-based interventions for depression.	Adults with depression (diagnosis or elevated symptoms) [Treatment]	QN	10	10	8	SG	For 10 RCTs: 60% completed ≥25% of modules 41% completed ≥50% of modules 30% completed ≥ 75% of modules 17% completed 100% of modules Predictors of completing <75%: male, younger, lower education, comorbid anxiety symptoms.
van Ballegooijen W, et al., 2014 [42]	RCTs of web-based interventions for depression	Adults with depression (diagnosis or elevated symptoms) [Treatment]	QN	24	11	11	G	% of completed sessions: 7 groups F2F CBT: 83.9%, CI: 75.7%-92.1% 7 groups iCBT: 80.8%, CI: 73.0%-88.7%. NSD between groups. % participants completed all sessions: 11 groups F2F CBT: 84.7%, CI: 78.0%-89.6%. 7 groups iCBT: 65.1%., CI: 55.3%-73.8%. Difference between groups significant (p<.001).

Table 2.5. Characteristics of systematic reviews examining adherence included in the current systematic review (continued).

Review	Included studies	Population [Study types[a]]	Analysis[b]	Studies in review (N)	Dep RCTs[c] (n)	Dep iCBT RCTs[d] (n)	Support types[e]	Adherence rates / conclusions of qualitative reviews
Baumeister H, et al., 2014 [47]	RCTs of Internet interventions, comparing guided vs unguided, guidance intensity or coach type.	Adults with mental disorder [Treatment]	QN	14	3	3	G vs SG	3 RCTs of guided vs unguided iCBT: Pooled mean no. of completed modules: SMD=0.50, CI: 0.19-0.82 (favours guided). Completer rates: OR: 1.60, CI: 0.70-3.65.
Cowpertwait L, et al., 2013 [51]	RCTs of web-based interventions for depression	Adults [Treatment]	QN	18	18	unclear	G, SG	18 RCTs: Attrition rate: M=0.33 (SD=0.15) Self-guided (n=7) larger attrition rate than guided (n=11): M=0.44, SD=0.09, β=0.19, p<.001
Richards D, et al., 2012 [52]	Studies of computer-based interventions for adults with depression	Adults with depression (diagnosis or elevated symptoms) [Treatment]	QN	45 (systematic review) 23 (meta-analysis: 19 trials)	19	17	G, SG	Treatment adherence analysed for 36 studies (includes non-depression): Failed to complete intervention: 57% OR for non-completion for different support: No support vs admin support: OR=2.45, p<.001 No support vs therapist: OR=7.35, p<.001 Admin support vs therapist: OR=3.0, p<.001

Table 2.5. Characteristics of systematic reviews examining adherence included in the current systematic review (continued).

Review	Included studies	Population [Study types[a]]	Analysis[b]	Studies in review (N)	Dep RCTs[c] (n)	Dep iCBT RCTs[d] (n)	Support types[e]	Adherence rates / conclusions of qualitative reviews
Christensen H, et al., 2009 [60]	RCTs of self-help Internet interventions for depression and anxiety, with non-active control.	No restriction [Mixed]	QL	23	8	Unclear	SG	% participants completed all modules: For 8 RCTs, ranged from 50-70%. Predictors of adherence: lower baseline depression, younger age, poorer depression literacy.

[a] *Type of studies included, based on intervention or population criteria: Treatment, Indicated (prevention), Selective (prevention), Universal (prevention), Mixed (treatment and prevention), Mixed prevention (indicated, selective and universal prevention).*
[b] *QN= Quantitative meta-analysis, QL= Qualitative review.*
[c] *Number of RCTs examining a depression intervention which uses technology.*
[d] *Number of RCTs examining a depression iCBT intervention.*
[e] *G= Guided support (either therapist or administrative support), SG=Self-guided.*
iCBT=Internet-delivered cognitive behaviour therapy; RCT=Randomised Controlled Trial; OR= Odds Ratio; CI= 95% Confidence Interval; F2F=Face-to-face; CMDs: Common mental disorders; M=Mean; SD= Standard deviation; NSD= No significant difference.

(a) Reviews of treatment studies (for depression diagnosis or elevated symptoms)

In the first of five reviews of treatment studies restricted to adults with depression (diagnosis or elevated symptoms), Richards and Richardson analysed data from 36 studies which reported adherence to a computerised or Internet treatment intervention and found that 57.0% of participants failed to complete the treatment intervention [52]. Non-completion of therapist supported interventions (28.0%) was less than for those with non-clinical 'administrative' support (38.4%). The rate of failure to complete was highest from interventions with no support (74.0%), and the odds ratio of non-completion of a self-guided intervention was 7.35 compared to therapist supported ($p < .001$) and 2.45 compared to an intervention supported by a non-clinician (p. 336). However, this 2012 Richards and Richardson review was not limited to iCBT interventions. Rather it included any computer-based psychological treatments for depression, including computer-based (non-Internet) CBT and problem-solving therapy alone.

In their 2013 review, Cowpertwait and Clarke undertook an analysis of 'drop-out rates' or 'attrition rates' [51]. The review did not define these terms or indicate how they were measured; however, since the analysis of drop-out was confined to treatment groups, we have assumed here that they refer to rates of adherence to the intervention. The review included 18 RCTs of web-based interventions for depression and reported a mean attrition rate of 0.33 ($SD = 0.15$), although it is not clear how many of the interventions trialled involved CBT. The review included guided and unguided studies. A moderator analysis found higher attrition rates for self-guided interventions ($n = 7$ trials, $M = 0.44$, $SD = 0.09$, $\beta = 0.19$, $p < .001$) compared to guided interventions ($n = 11$ trials).

van Ballegooijen and colleagues examined and compared intervention adherence in RCTs of iCBT with adherence in RCTs of face-to-face CBT [42]. To be included in the review, a study must have incorporated at least one group that was administered either face-to-face CBT or guided iCBT, included participants who were largely or entirely recruited from the community, and have been published after 2000. The review identified 24 trials comprising 26 treatment groups. No study employed a direct comparison between adherence in a face-to-face and iCBT group. The percentage of the intervention undertaken by participants was compared for the seven groups of participants receiving face-to-face CBT (83.9% of sessions completed, 95% CI: 75.7%, 92.1%) and the eight groups of participants receiving iCBT (80.8% of sessions completed, 95% CI: 73.0%, 88.7%). No significant difference in attrition rates was detected between the groups. However, there was a significant difference between the number of participants who completed the entire intervention with participants receiving face-to-face CBT more likely to finish the program (84.7%, 95% CI: 78.0%, 89.6%) than those receiving iCBT (65.1%, 95% CI: 55.3%, 73.8%).

Baumeister et al. published a quantitative meta-analytic review of adherence outcomes in treatment trials which directly compared adherence for unguided- relative to guided-iCBT for mental disorders in adults and identified three trials targeting depression [47]. The authors reported that participants assigned to guided iCBT for depression completed significantly more modules than those in unguided iCBT conditions (*SMD* = 0.50, 95% CI: 0.19, 0.82), and were more likely to complete the whole iCBT program (OR: 1.60, 95% CI: 0.70, 3.65).

Karyotaki et al. restricted their review to a consideration of adherence among participants in self-guided web-based interventions [37]. In their individual patient data meta-analysis of data from 10 RCTs, Karyotaki and colleagues found that only 17% of participants completed all intervention modules, 30% completed 75% or more of the modules, and 41% completed half or more of the assigned intervention modules. A large proportion of participants (40%) engaged with less than one-quarter of the intervention. Lack of adherence was predicted by being male, younger, lower educational attainment, and having comorbid anxiety symptoms.

Overall the reviews suggest that adherence is higher for guided than unguided treatments and that participants are more likely to finish a face-to-face program of CBT than to complete a guided iCBT program. The overall completion rate for unguided and guided interventions ranged from 17.0% to 26.0% (where reported) and 61.6% to 72.0%, respectively. Reviews that reported odds ratios found that participants were between 1.60 and 7.35 times more likely to complete a guided compared to an unguided intervention.

(b) Reviews of intervention studies (no symptom restriction)

The earliest systematic review of adherence to Internet interventions was published in 2009 by Christensen, Griffiths and Farrer, and included eight RCTs of self-guided interventions for depression. However, these interventions were not limited to iCBT, including any self-help website for depression or anxiety disorder [60]. The review did not exclude trials based on user characteristics or symptom levels. The identified adherence outcome data for the included trials were varied and included number of logins, number of modules or sessions completed, time spent on the intervention, and number of postings on a discussion board. For those trials for which relevant completion data was reported, the number of users completing the intervention ranged from 50 to 70% of users. Christensen et al. also found that increased adherence was associated with lower levels of baseline depression, younger age and poorer depression literacy.

The other review which examined adherence to depression iCBT interventions, and which did not restrict selection by symptoms, examined RCTs of Internet interventions that included one or more

gamification features and which targeted common mental disorders [34]. The review included 30 RCTs of depression interventions using technology. Unfortunately, the primary meta-analysis in the study was limited to participant "adherence to the study protocol", measured as a percentage of those participants who completed post-intervention and follow-up assessments. This concept is not the focus of the present review. However, the authors did report some usage data, a measure which is consistent with the definition of 'adherence' as used in this review. Of the 61 RCTs in the full review (i.e. not limited to those for depression), 34 reported on the program completion rate, with an average of 54.0% of participants completing the intervention (*SD* = 24.6%). Usage data was not reported for the subset of depression interventions identified in the review.

2.3.6 Cost-effectiveness of iCBT for depression

Cost-effectiveness analyses evaluate the costs and benefits of a particular treatment from a specific perspective. For example, an analysis which takes a societal perspective incorporates the costs and benefits to society as a whole, whereas one which takes a healthcare provider perspective considers costs and benefits relevant to that organisation only [69]. Such analyses can, therefore, be tailored to the needs of different stakeholders. Most commonly the cost-effectiveness of interventions is reported as an *Incremental Cost-Effectiveness Ratio (ICER)* – a ratio of the extra cost of providing the intervention (relative to standard care or a different intervention) to the extra health benefit achieved (relative to standard care or a different intervention) [70, 71]. For example, the ICER may be calculated to reflect the cost per 'Life Years Gained' (LYGs). *Cost-utility analysis* (a subset of cost-effectiveness analysis) takes into account the patient's quality of life (not just length of life), for example through the use of 'Quality-Adjusted Life Years Saved' (QALYs). Thus, a cost-utility ICER would reflect the cost of each QALY gained through the use of the intervention compared to standard care (or a different intervention) [72]. ICERs are frequently used by policy makers to make funding and implementation decisions. For example, the National Institute of Clinical Excellence (NICE) in the United Kingdom compares technology interventions based on the 'cost per QALY gained' ICER, and generally recommends that an ICER of £20,000 per QALY gained is acceptable [73].

Reviews of studies

The current study identified three systematic reviews of the cost-effectiveness of iCBT for depression [39, 48, 54] which are summarised in Table 2.6. In the most recent of these, Donker and colleagues published a 2015 review which focused on economic evaluations of Internet-interventions for mental health outcomes relative to control, without any restriction on the trial population [39]. Of the 16 trials included in the study, four were studies of depression iCBT. The results of these four studies are described below. The other two reviews comprised subsets of the trials reported in the 2015 review.

Table 2.6. Characteristics of systematic reviews examining cost-effectiveness included in the current systematic review.

Review	Included studies	Population [Study types[a]]	Studies in review (N)	Dep RCTs[a] (n)	Dep iCBT RCTs[b] (n)	Support types[c]	Cost-effectiveness conclusions
Donker T, et al., 2015 [39]	RCTs of economic evaluations of interventions for mental health symptoms	No restriction [Treatment, Indicated]	16	4	4	G, SG	4 studies of cost-effectiveness for depression: a) Netherlands, societal perspective. G iCBT: 30% probability of being cost-effective at WTP=0. iCBT vs WLC: Cost-utility CER= €22,609. b) Netherlands, societal perspective SG iCBT 60% probability of being cost-effective at WTP=0. c) U.K., health service provider perspective G iCBT better outcomes but at a higher cost than TAU. iCBT vs TAU: ICER = £3,528 iCBT vs TAU: Cost-utility CER= £17,173 d) U.K., societal perspective (workplace setting) No differences outcomes for iCBT vs attention control. Cost of lost employment and absence from work higher for control group at follow-up, but not sig. (£111 vs £96, p=.76; £143 vs £119, p=.64).
Arnberg FK, et al., 2014 [48]	RCTs of Internet interventions for anxiety and mood disorders	All ages with diagnosis of anxiety or mood disorder [Treatment]	2	1	1	G, SG	1 study of cost-effectiveness for depression a) U.K., health service provider perspective G iCBT better outcomes but at a higher cost than TAU. iCBT vs TAU: ICER = £3,528 iCBT vs TAU: Cost-utility CER= £17,173

Table 2.6. Characteristics of systematic reviews examining cost-effectiveness included in the current systematic review (continued).

Review	Included studies	Population [Study types[a]]	Studies in review (N)	Dep RCTs[a] (n)	Dep iCBT RCTs[b] (n)	Support types[c]	Cost-effectiveness conclusions
Hedman E, et al., 2012 [54]	RCTs of iCBT for a clinical problem	Adults [Treatment, Indicated]	8	3	3	G, SG	3 studies of cost-effectiveness for depression: a) Netherlands, societal perspective. G iCBT 30% probability of being cost-effective at WTP=0. iCBT vs WLC: Cost-utility CER= €22,609. b) Netherlands, societal perspective SG iCBT 60% probability of being cost-effective at WTP=0. c) U.K., health service provider perspective G iCBT better outcomes but at a higher cost than TAU. iCBT vs TAU: ICER = £3,528 iCBT vs TAU: Cost-utility CER= £17,173

[a] Type of studies included, based on intervention or population criteria: Treatment, Indicated (prevention), Selective (prevention), Universal (prevention), Mixed (treatment and prevention), Mixed prevention (indicated, selective and universal prevention).

[b] Number of RCTs examining a depression intervention which uses technology.

[c] Number of RCTs examining a depression iCBT intervention.

[d] G= Guided support (either therapist or administrative support), SG=Self-guided

iCBT=Internet-delivered cognitive behaviour therapy; RCT=Randomised Controlled Trial; WLC=Wait-list control; WTP= Willingness to pay for additional improvement; TAU= Treatment as usual; ICER= Incremental cost-effectiveness ratio; CER=Cost effectiveness ratio.

Arnberg et al. restricted their study selection to RCTs of Internet interventions for participants of any age with a diagnosis of anxiety or depressive disorder and identified one study [48]. Hedman and colleagues focused on RCTs of iCBT for clinical problems in adults (without symptom restriction) and included three studies [54].

The first economic evaluation, published in 2010, and included in all three reviews, was a study of psychologist-guided iCBT conducted in the United Kingdom. The study, which took a healthcare provider perspective [74], found the iCBT intervention yielded better outcomes than treatment as usual, but at a higher cost. The cost per QALY gained through the use of the guided iCBT intervention was £17,173, and the ICER for the cost per extra patient recovering was £3,528. However, the paper did not provide a definition of a recovered patient.

The Donker and Hedman reviews also each included two studies of the cost-effectiveness of iCBT conducted in the Netherlands, both of which were undertaken from a societal perspective [75, 76]. First, a study of guided iCBT in adults with depression symptoms found that the cost per QALY gained was €22,609 compared to a wait-list control, and that the ICER for the cost for a reliably improved patient was €1,817 (the cost for a clinically significant change in depressive symptoms using the iCBT intervention rather than a wait-list) [75]. This study also found that if society had no willingness to pay for one extra QALY, the probability that iCBT is superior to a wait-list control (more health benefits for lower costs) is 28%. Second, a study of unguided iCBT in adults with at least mild symptoms of depression found that if society had no willingness to pay for an extra QALY, the unguided iCBT intervention had a 60% probability of being cost-effective compared to treatment as usual [76]. If the willingness to pay was increased to €80,000, then this probability decreased to 40% compared to treatment as usual. This paper noted that €18,000 is the generally accepted threshold for cost per QALY for preventative care in the Netherlands (p. 317).

The final study of cost-effectiveness, included only in the Donker review, examined the use of unguided iCBT compared to a psycho-educational control, in a workplace setting [77]. The trial found no difference in depression outcomes for the intervention group compared to the control group, and no differences in QALYs gained. However, the control which comprised website pages about mental health was potentially an active condition (since psycho-education has previously been shown to be effective in reducing depression symptoms [78]), and the depression scores decreased for both groups at follow-up. The costs of lost employment and absence from work at follow-up were higher for the control group compared to the intervention group, but these differences were not significant.

2.4 Discussion

The current meta-review identified 46 systematic reviews of trials of the effectiveness or cost-effectiveness of iCBT for depression, and overall the reviews were of high quality when assessed against an adapted quality assessment tool for systematic reviews (AMSTAR). Only a minority of reviews focused on the use of iCBT for prevention, possibly due to the relatively smaller number of RCTs that have been undertaken in this domain.

2.4.1 Effectiveness of iCBT in the treatment and prevention of depression

Overall, it was found that iCBT is an effective treatment intervention across age groups with overall effect sizes from eight quantitative meta-analyses of adults ranging from small (0.33) to large (0.90) and a meta-analysis of four treatment studies in youth aged 25 years and younger yielding a moderate effect (*SMD* = 0.76, 95% CI: 0.41, 1.12). Similarly, the four reviews of unrestricted populations (adults and youth) reported positive results. However, the status of iCBT for the prevention of depression was less clear, particularly with respect to the type of prevention programs that might be effective. There was evidence of the effectiveness of iCBT as an indicated prevention tool for adults from one meta-analysis (*SMD* = 0.46, 95% CI: 0.22, 0.70, *n* = 9 RCTs [30]) but there was less evidence for the use of iCBT for universal prevention in adults, with the only review of this topic including only three prevention trials of Internet interventions for depression, one of which was a mindfulness-based program [31]. Similarly, there was evidence in support of the effectiveness of iCBT for preventing depression among youth, with the sole review of the topic reporting positive results for five out of six included RCTs and significant depression reductions in male participants only for the sixth RCT [40]. However, the review investigated iCBT depression prevention programs for youth regardless of the type of prevention. Finally, there was some evidence from mixed intervention reviews that iCBT may be effective for prevention and treating symptoms of women in the perinatal period.

The evidence in support of iCBT for the treatment of depression or the reduction of depressive symptoms is compelling. However, it is clear that further research is required to investigate the effectiveness of iCBT in selective and universal prevention in adults. Further, research is also required to investigate the effectiveness of different types of preventive intervention among youth. This population is particularly important because depression typically first appears during adolescence or up to 25 years old [79], the rate of recurrence of depression is high [80, 81] and intervention using scalable methods has the potential to avert substantial level of disability over the individual's lifetime.

2.4.2 The effect of human guidance

The question of whether human guidance is necessary or generates superior outcomes when used in implementing iCBT programs is critical given its resource implications and potential impact on cost-effectiveness (see below). It is, however, important to note that there are potential negative effects of iCBT for depression, and that these risks may be greater for unguided compared to guided interventions. For example, users may not experience any improvement in their symptoms [82], users may deteriorate rather than improve during their use of iCBT (i.e. their depression symptoms increase) [83], or users may become distressed in response to program content. Human guidance, particularly through monitoring users' symptoms and treatment progress, may minimise the impact of such risks and allow for identification of individuals for whom iCBT may not be appropriate.

Many treatment reviews considered the impact of human guidance on treatment effectiveness, with mixed results. Although earlier analyses reported significant differences between guided and unguided iCBT, more recent meta-analyses did not find a significant difference in effectiveness between trials of each type of intervention.

It should be noted that there are limitations associated with drawing conclusions for the comparative effectiveness of guided versus unguided interventions from pooled results, given that the included trials varied on a number of parameters. For example, the trials were conducted in a range of primary care, community, university and school settings, and iCBT interventions varied in length, content, and where guided, the type and level of human support. The potential for outcome improvement in a guided treatment trial of iCBT among clinic-diagnosed participants may differ from that in an unguided prevention trial in a universal sample of university students. In order to more rigorously compare the impact of guidance, 'head-to-head' trials (of participants selected from the same population, setting and intervention, but with and without guidance) are required. A possible limitation of head-to-head trials of guided versus non-guided iCBT is that such trials may require that participants be willing to be randomised to face-to-face guidance. This may mitigate against the recruitment of individuals who have a preference for Internet interventions and for whom arguably such interventions might offer the most benefit. Nevertheless, such trials do provide an opportunity for direct comparison of guided versus non-guided iCBT, and in many cases human support can be provided via technology (e.g., via email or messaging). One included review did focus on head-to-head trials and found a non-significant difference favouring guided over unguided interventions [47]. However, given the review was based on only three trials, further head-to-head research is required to clarify the relative effectiveness of guided and unguided treatment across a range of settings and participants.

Putting aside the relative effectiveness of guided and unguided iCBT, it is noteworthy that there is evidence that the latter does work for the treatment of depression. For example, a recent individual participant data analysis based on 13 treatment trials of exclusively non-clinician-supported iCBT calculated a small (albeit significant) effect size of 0.27 [28]. Since unguided iCBT can be fully automated, it is highly scalable, and so clinical effectiveness has the potential to translate into substantial public health gains.

2.4.3 Participant engagement

Participants' engagement with iCBT programs may also be important, with a study published in 2011 finding that module completion was correlated with symptom improvement in trials of Internet interventions for depression and anxiety [68]. Seven of the 46 reviews in the current study (15.2%) focused on intervention adherence and most commonly measured this as completion of the program. In the reviews of treatment studies, program completion was higher for guided compared to unguided CBT, and participants were less likely to complete guided iCBT compared to a course of face-to-face CBT. Unguided completion rates ranged from 17.0% to 26.0% whereas completion rates for guided CBT ranged from 61.6% to 72.0%. In addition, two reviews reported on completion rates for studies unrestricted by symptom levels – one which was published in 2009 and reported completion rates from 50% to 70%, and the other published in 2016 which reported that, on average in 34 trials, 54% of participants completed the intervention, although this included programs targeted at disorders other than depression.

The results of those reviews which focused on effectiveness, combined with those which focused on adherence, suggest that although fewer participants completed unguided iCBT interventions compared to those where there was human guidance, the effectiveness of both types of interventions could be similar for the treatment of depression. It is possible that users 'take what they need' from an unguided intervention [84] whereas human guidance engages them and creates motivation to complete the whole intervention.

2.4.4 Cost-effectiveness

From a health policy perspective, the relative effectiveness of unguided iCBT is important, since the ongoing and incremental cost of delivering iCBT without human guidance is likely to cost much less. Unfortunately, however, there are few studies of the cost-effectiveness of iCBT (either guided or unguided). The current study identified three reviews focusing on studies of cost-effectiveness, which comprised only four studies in total. Of the two studies that evaluated guided iCBT, one U.K. study took a healthcare provider perspective and reported the cost per QALY gained as £17,173 for

the guided intervention compared to treatment as usual, a figure which fell within the £20,000 threshold set by NICE for a cost-effective intervention. The other study, conducted in the Netherlands, took a societal perspective and calculated the cost per QALY gained as €22,609. Of the two studies of cost-effectiveness that evaluated unguided iCBT, one conducted in the Netherlands found that iCBT had a 60% chance of being cost-effective compared to treatment as usual if society had no willingness to pay for an extra QALY. The other study was conducted in a workplace in the U.K. and did not find a significant intervention effect for depression.

This limited number of studies suggests that, given there is a larger body of evidence that iCBT is effective for the treatment of depression, it is highly likely that both guided iCBT and unguided iCBT are cost effective when compared to health intervention thresholds. However, more studies of cost-effectiveness are required.

2.4.5 Limitations

There are a number of limitations associated with the current study and the reviews it identified. First, there is large variability in the reviews and the studies that they include. Some reviews were limited to CBT-based Internet interventions, whereas others included interventions based on a range of (non-CBT) psychological approaches. Similarly, not all reviews were restricted to Internet-delivered interventions, or to interventions specifically targeted at depression. There were also differences in criteria for determination of diagnosis status (or elevated symptoms) of study populations, including the use of different depression measures. Further, there is an inherent risk of publication bias in any systematic review, and this was not assessed by 60.9% of the reviews. The included reviews also included a large range of studies with varying characteristics, including different populations, settings, outcome measures and interventions. Many of the quantitative reviews pooled results from trials with a high level of heterogeneity.

Finally, the studies in this meta-review were identified and selected for inclusion by a single reviewer who also performed the data extraction. Although not best practice for undertaking a systematic review this was necessary for pragmatic reasons (post-graduate thesis).

2.5 Conclusion

The systematic reviews included in the current review provide evidence that iCBT interventions are generally effective for the treatment of depression and depression symptoms in adults and young people. However, the evidence regarding the effectiveness of iCBT for the prevention of depression is limited by the small number of published studies. Similarly, although the included reviews

(particularly the most recent reviews) suggest that there may not be a significant difference in the effectiveness of guided and unguided interventions, since few head-to-head trials are included, the question of whether human guidance results in superior outcomes is not adequately answered.

To better understand the potential of iCBT for depression more reviews (or sub-reviews) of focused trials need to be conducted, for example with stricter inclusion and exclusion criteria. This would be possible if there were more research trials which focus on different types of prevention, and more head-to-head trials investigating the role of varying levels of human guidance within the same population and setting. In addition, reviews which explore the heterogeneity between studies could attempt to identify the most effective intervention components or trial designs.

The cost-effectiveness of iCBT also requires further research attention. Although there is some evidence that iCBT is cost-effective compared to treatment as usual, more cost-effectiveness analyses need to be conducted, particularly those which integrate questions of the relative importance of human guidance. This evidence could then inform not only the integration of iCBT into health care systems (with guidance as indicated) but also the potential and feasibility of unguided, population-wide iCBT for public health benefits.

3. Unguided automated iCBT for depression: A systematic qualitative review

3.1 Introduction

The systematic review in Chapter 2 demonstrated the effectiveness of iCBT for the treatment of depression. Further, the review provided broad evidence that unguided iCBT interventions (i.e., where no human-based therapeutic support is provided to the user) can be effective for depression. Given the potential of unguided programs to be delivered en-masse at low cost, particularly if human input is not required, their effectiveness is of substantial public health relevance. However, current reviews of unguided iCBT are the subject of several limitations (see below). Accordingly, this chapter reports the outcome of a systematic review of the effectiveness of automated iCBT interventions for depression.

3.1.1 The importance of automated iCBT interventions

As noted in Chapter 1, once developed, unguided iCBT interventions can be delivered at a relatively small cost compared to interventions that require human input. The scalability and global reach of technology present an opportunity to deliver unguided iCBT to large numbers of people worldwide, potentially providing an effective psychological intervention to an unlimited number of individuals [12]. Given the high prevalence of depression internationally [1, 85] and that depression is one of the leading causes of disability [86], providing widespread access to an effective treatment resource could produce a significant public health impact.

Scalability of treatment delivery can be maximised if unguided iCBT interventions are fully automated. An automated program is one that does not employ direct human input in its delivery. It includes neither therapist input nor more minimal levels of contact or guidance from non-therapists such as coaches, teachers or other guides. Rather, use of an automated iCBT intervention is completely self-directed by the user, with prompts, feedback or other messaging automatically generated by the technology system, based on inputs from the user. As such, automated iCBT interventions can be considered a subset of self-guided or unguided interventions.

3.1.2 The extant literature on the effectiveness of automated iCBT interventions

To better understand the potential of automated iCBT for the treatment of depression, it is instructive to specifically examine the effectiveness of interventions of this type. Several existing

reviews provide helpful analysis. However, there are limitations associated with these reviews for the current purpose.

One meta-analysis focused exclusively on unguided iCBT for depression but included trials of intervention conditions that would not be classified as automated because support was provided by a non-therapist [28]. For example, the review included an RCT of the effectiveness of the *MoodGYM* iCBT program where participants received a weekly phone call from a health coach [87].

Two quantitative reviews have categorised and separately examined trials of computer-based interventions depending on the level and type of treatment support provided, including those that were automated [51, 52]. In their 2012 review of computer-based interventions for depression in adults, Richards and Richardson [52] identified 12 studies where the intervention delivery was completely unsupported and reported a mean post-treatment effect size of 0.58 (95% CI: 0.28, 0.88). However, this review included interventions that were not based on CBT, interventions which were computer-based but not delivered over the Internet (for example, delivered via CD-ROM), and studies that were not RCTs [52]. Cowpertwait and Clarke reviewed RCTs of web-based interventions for depression in adults in their 2013 paper but included interventions that were not CBT-based [51]. They identified five trials that did not include human-support and reported a Hedges *g* effect size for depression at post-intervention for these trials of 0.29 (95% CI: 0.19, 0.39).

3.1.3 The current study

Given the limitations associated with these existing reviews, the current study is an updated systematic qualitative review of RCTs of unguided, automated iCBT for depression. Unlike the previous reviews, the current study is limited to automated interventions based on CBT and delivered over the Internet. Except for the exclusion of studies targeting those with a physical illness or a mental disorder other than depression, this review places no restriction on study populations. Studies of populations of any age range, diagnostic status or symptom severity or from any setting are eligible for inclusion. A quantitative meta-analysis was not undertaken due to this heterogeneity [88, 89] and a qualitative systematic review was considered the most appropriate method for describing the current state of the literature.

3.2 Method

The current review was guided by the *Preferred Reporting Items for Systematic Reviews and Meta-Analysis 'PRISMA'* statement [21]. All searches and assessments were undertaken by a single reviewer (the candidate).

3.2.1 Inclusion and exclusion criteria for systematic review

Studies were eligible for inclusion if they:

i) reported on a randomised controlled trial (RCT) which:

   a) contained at least one condition where participants received a fully automated iCBT intervention;

   b) contained at least one control comparison condition; and

   c) reported a depression outcome based on a standard measure.

ii) were published in a peer-reviewed, English-language journal; and

iii) were published after January 2000 and before June 2018.

Studies were excluded if the:

i) trial participants were selected because they had a physical health disorder or a mental health condition other than depression;

ii) iCBT intervention was designed to treat or prevent a physical or mental health condition other than depression;

iii) iCBT intervention was computerised, but not delivered over the Internet (e.g. delivered via CD-ROM or via a mobile or smartphone app);

iv) iCBT intervention did not predominantly provide training in CBT; and

v) iCBT intervention condition included any human contact during the intervention delivery period (e.g., emails, chat, phone calls or postcards, from a therapist, coach, teacher, lay person or study personnel).

Studies were excluded if they were non-inferiority trials or randomised controlled pilot studies (as defined by Eldridge and colleagues [90]). In addition, interventions that only utilised one CBT component (e.g., structured problem-solving) were excluded (e.g., [91]).

3.2.2 Search strategy

Studies were identified using two approaches. First, all potentially relevant studies were extracted from each of the systematic reviews identified in Chapter 2. Second, in order to ensure that more recently published studies were included in the current review, a systematic search of articles published from 2012 was undertaken. In July 2018, the PubMed, Cochrane Library and PsycINFO databases were searched electronically to identify controlled trials based on the intersection of the three key concepts 'depression', 'Internet', and 'cognitive behavioural therapy' using the search terms listed in Table 3.1.

Table 3.1. Search strategy for systematic review

PubMed search terms	
1	(depression OR major depressive disorder OR depressive) [All fields]
2	(online delivery OR internet OR web based OR web) [All fields]
3	(cognitive behavioral therapy OR cognitive behavioural therapy OR behavioral therapy OR behavioural therapy OR cognitive therapy OR iCBT OR CBT) [All fields]
4	(randomized OR randomised OR placebo OR randomly OR trial) [All fields]
5	(controlled clinical trial [Publication Type]) OR randomized controlled trial [Publication Type])
6	1 AND 2 AND 3 AND 4 AND 5 with filter: Publication date 1/1/2012 - 30/6/2018
Cochrane Library search terms	
1	(depression OR major depressive disorder OR depressive) in Title, Abstract, Keywords
2	(online delivery OR internet OR web based OR web OR online) in Title, Abstract, Keywords
3	(cognitive behavioral therapy OR cognitive behavioural therapy OR behavioral therapy OR behavioural therapy OR cognitive therapy OR iCBT OR CBT) in Title, Abstract, Keywords
4	1 AND 2 AND 3 with filter: Publication year 2012 - 2018
PsycINFO search terms	
1	(depressive OR depression).mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures]
2	(internet OR web OR online OR web-delivered).m_titl.
3	(cognitive OR behavioural OR behavioral OR iCBT OR CBT).mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures]
4	(RCT OR randomised OR randomized OR randomly OR placebo).m_titl.
5	1 AND 2 AND 3 AND 4 with filter: Publication year 2012 - 2018

3.2.3 *Data extraction*

Records identified in the database searches were downloaded into referencing software (EndNote X8) for abstract screening. Following screening, the full-text versions of potentially relevant articles were retrieved for further assessment. Additionally, the full-text of articles identified through systematic reviews was downloaded and checked for eligibility.

The following data were then extracted from eligible studies and recorded in an Excel spreadsheet:

- study setting and country;
- participant inclusion/exclusion criteria;
- number of randomised participants;
- details of age/gender (where gender or mean age were reported separately for each condition, data were combined);
- trial condition groups;
- automated iCBT intervention details;
- depression outcome measure and effectiveness results;
- details of any outcome predictor analyses; and
- details of participant adherence to the automated iCBT condition.

The between-group effect size (Cohen's *d* or Hedge's *g*) for the automated iCBT versus the control group were recorded if reported or calculated if possible. These effect sizes were computed as the difference between the means for the control and the intervention conditions at post-intervention, divided by the pooled standard deviation at post-test (M_C-M_I)/*SD*. Estimated means from intention to treat (ITT) analyses were used, adjusted for other variables such as baseline symptom severity if these data were reported. For non-ITT analyses, or where calculation of the effect size was not possible based on the reported data, available results were recorded.

3.2.4 *Quality of studies*

The quality of the included studies was assessed using the Cochrane Collaboration's tool for assessing risk of bias in randomised trials [25]. Replicating the method used by Ebert and colleagues in their 2016 meta-analysis of Internet-based guided-self-help for depression [33], four criteria were considered. These were whether:

i) there was an adequate generation of the condition allocation sequence (selection bias);

ii) allocation of participants to conditions was conducted by an independent party (selection bias);

iii) outcome assessors were blinded (performance bias); and

iv) the analysis was conducted on an intention to treat basis, with all randomised participant data included (attrition bias). Performance bias was assessed as being handled adequately if assessors were blinded to the participant condition, or if the outcomes were self-reported by participants.

3.3 Results

3.3.1 Search results

The systematic search of PubMed, Cochrane Library and PsycINFO databases resulted in 746 citations with a publication date from 2012 (see the study selection flow diagram in Figure 3.1). Of these, 246 were identified as duplicates and removed. Abstracts of the remaining citations were screened against the inclusion and exclusion criteria, and a further 444 records were excluded. The full text of the remaining 56 articles was obtained and assessed for eligibility. Of these, 46 were assessed as ineligible for inclusion. The main reason for exclusion was that the iCBT condition contained some kind of human contact or support during the intervention period (*n* = 35, 77.8%). This support included phone-calls, emails, electronic chat, postcards and teacher supervision, and one article was assessed as ineligible since, as an ethics approval requirement, all participants were emailed once during the intervention period to check for problems or concerns about participation [10]. Thus, 10 eligible articles were identified from the database search. In addition, 11 eligible papers were identified from previously published systematic reviews. After removing duplicates, a total of 19 articles reporting on 17 RCTs, were eligible for inclusion.

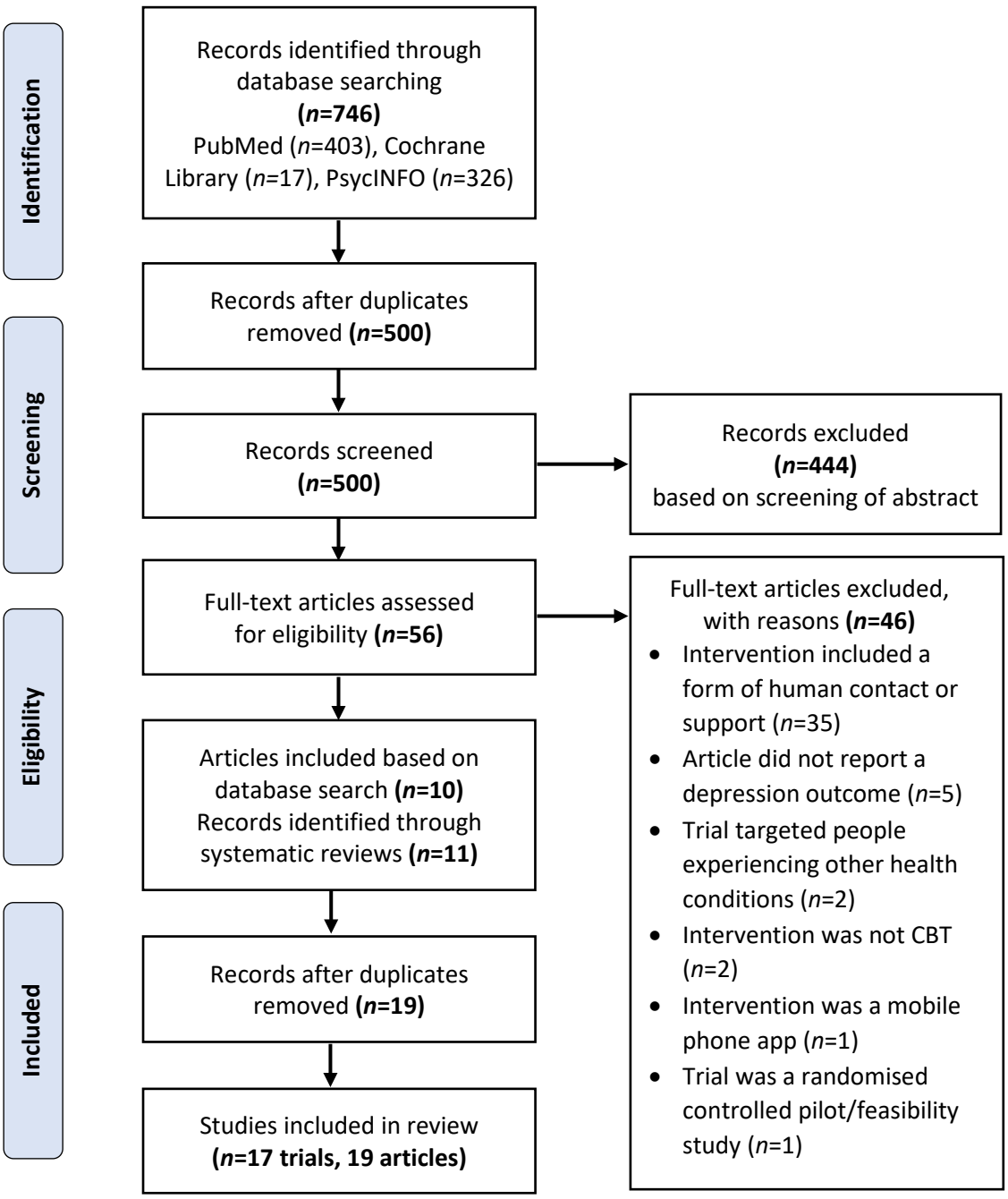

Figure 3.1. Preferred reporting items for systematic reviews and meta-analysis (PRISMA) diagram displaying procedure for selecting articles

3.3.2 Study characteristics

The characteristics of the included studies are summarised in reverse chronological order in Table 3.2 and are described below.

Table 3.2. Characteristics of studies included in the current systematic review.

Study **Intervention (language)**	**Country** **Recruitment setting**	**Inclusion criteria** **Exclusion criteria**	**N** **% female** **Mean age (SD)**	**Trial conditions** [a]	**Depression measure(s)** **Assessment timepoints**	**Depression outcome iCBT vs control** [b] **Effect size (95% CI)**	**Intervention adherence**
Mira A, et al. 2017 [92] Smiling is Fun (Spanish)	Spain Community (Internet and print advertising)	Adults 18-65 years with current stressful event, BDI-II ≤ 28 Exclusion: current treatment, treatment in last year, severe Axis I mental disorder, suicidal ideation.	124 67.7% female Age: M=35.6 (SD=9.70)	1. iCBT with automated support (**iCBT**) n=44 2. iCBT + human support (Guided) n=44 3. Wait-list control (**WL**) n=44	BDI-II 12 weeks 12 months~	<u>BDI-II at post-test (12 weeks)</u> iCBT vs WL: d=0.50 (0.05; 0.94)	iCBT group: 77.77% completed all modules.
Noguchi R, et al. 2017 [93] Brief, simplified iCBT program (Japanese)	Japan Community (marketing company website)	Adults 19-66 years, at least mild depression (CES-D ≥ 16 and PHQ-9 ≥ 5) Exclusion: suicidality.	974 50.1% female Age: M=43.7 (SD=11.3)	1. i**CBT** n=326 2. Simplified Emotion-focused mindfulness n=323 3. Wait-list control (**WL**) n=325	CES-D PHQ-9 6 weeks 12 weeks 18 weeks~	*For all participants:* <u>CES-D, PHQ-9 at post-test (6 weeks)</u> iCBT vs WL: NSD <u>CES-D, PHQ-9 at 12 weeks</u> iCBT vs WL: NSD *Those who met inclusion criteria:* <u>CES-D at post-test (6 weeks)</u> iCBT vs WL: d=0.34[†] (0.16, 0.51)[†] <u>PHQ-9 at post-test (6 weeks)</u> iCBT vs WL: d=0.27[†] (0.10, 0.44)[†] <u>CES-D, PHQ-9 at 12 weeks</u> iCBT vs WL: NSD	Not reported

Table 3.2. Characteristics of studies included in the current systematic review (continued).

Study **Intervention (language)**	**Country** **Recruitment setting**	**Inclusion criteria** **Exclusion criteria**	**N** **% female** **Mean age (SD)**	**Trial conditions** [a]	**Depression measure(s)** **Assessment timepoints**	**Depression outcome iCBT vs control** [b] **Effect size (95% CI)**	**Intervention adherence**
Lokman S, et al. 2017 [94] Complaint-directed mini-interventions (CDMIs) (Dutch): Sleep, stress, worry	Netherlands Community (websites, social media, online newsletters)	Adults ≥ 18 years with mild to moderate depressive symptoms, 14 ≤ IDS-SR ≤ 38 Exclusion: suicidality	329 75.7% female Age: M=43 (SD=12.93)	1. CDMIs (**iCBT**) n=165 2. Wait-list control (**WL**) n=164 iCBT participants: Sleep: 59 (35.8%); Stress: 45 (27.3%); Worry: 61 (37.0%)	IDS-SR 3 months 6 months˜	Model adjusted for age, gender, living arrangement, education, depression and anxiety scores at baseline. <u>IDS-SR at post-test (3 months)</u> iCBT s vs WL: d=0.39† (0.17, 0.61)†	iCBT group: 127 (77.0%) created account, 91.3% accessed at least once. Median number log-ins=3 (range 0-166).
Silverstone P, et al. 2017 [95] *MoodGYM* (English)	Canada Primary care	Adults ≥ 18 years	1,489 Gender/age not reported	1. Treatment as usual (**TAU**) n=432 2. TAU plus screening n=426 3. Screening, TAU plus iCBT (**iCBT**) n=440 4. Screening, stepped care n=191	PHQ-9 12 weeks	*Completer analysis only (no ITT)* <u>PHQ-9 at 12 weeks</u> iCBT vs TAU: All completers (n=530): NSD Depressed only (PHQ-9 ≥ 10) (n=72): NSD	Of 29 depressed completers in iCBT group: 25 logged on, <5 completed more than 1 session, none completed whole program.

Table 3.2. Characteristics of studies included in the current systematic review (continued).

Study **Intervention (language)**	**Country** **Recruitment setting**	**Inclusion criteria** **Exclusion criteria**	**N** **% female** **Mean age (SD)**	**Trial conditions** [a]	**Depression measure(s)** **Assessment timepoints**	**Depression outcome iCBT vs control** [b] **Effect size (95% CI)**	**Intervention adherence**
Montero-Marin J, et al. 2016 [96, 97] *Smiling is Fun* (Spanish)	Spain Primary care	Adults 18-65 years, mild to moderate severity symptoms (14-28 on BDI-II), depression diagnosis on MINI interview. Exclusion: severe psychiatric disorder, severe depression (BDI-II ≥ 29)	296 75.7% female Age: M=42.9 (SD=10.33)	1. 'Improved' Treatment as usual by GP (**iTAU**) n=102 2. TAU + iCBT (**iCBT**) n=98 3. TAU + guided iCBT (GiCBT) n=96	BDI-II 3 months 6 months 12 months	*ITT multi-level analysis for repeated measures, calculating regression coefficients (adjusted for baseline scores):* <u>BDI-II at post-test (3 months)</u> iCBT vs iTAU: NSD <u>BDI-II at 6 months</u> iCBT vs iTAU: B=2.91, p<.001 in favour of iCBT <u>BDI-II at 12 months</u> iCBT vs iTAU: B=3.69, p<.001 in favour of iCBT	iCBT group: At 15 months, 72.4% had accessed iCBT, 41.8% completed >6 modules. Mean no. completed modules=4.

Table 3.2. Characteristics of studies included in the current systematic review (continued).

Study **Intervention (language)**	**Country** **Recruitment setting**	**Inclusion criteria** **Exclusion criteria**	**N** **% female** **Mean age (SD)**	**Trial conditions** [a]	**Depression measure(s)** **Assessment timepoints**	**Depression outcome iCBT vs control** [b] **Effect size (95% CI)**	**Intervention adherence**
Meyer B, et al. 2015 [98] *Deprexis* (German)	Germany Community and clinics	Adults 18-65 years, PHQ-9 ≥ 15, depression diagnosis on MINI interview. Exclusion: suicidality, schizophrenia, bipolar disorder.	163 74.8% female Age: M=42 (SD=11.39)	1. **TAU** n=78 2. TAU + iCBT for 3 months (**iCBT**) n=85	PHQ-9 3 months 6 months	*Based on observed means* <u>PHQ-9 at post-test (3 months)</u> iCBT vs TAU: d=0.57 (0.22, 0.92) <u>PHQ-9 at 6 months</u> iCBT vs TAU: d=0.33 (-0.03, 0.69) No effects of initial depression severity, gender, age, educational status, concurrent psychotherapy, onset age of depression. Concurrent anti-depressant use - significant interaction effect.	iCBT group: 56 (71.8%) completed at least 4 sessions and at least 60 minutes.
Twomey C, et al. 2014 [99] *MoodGYM* (English)	Ireland Secondary care (waiting lists for mental health services)	Adults ≥ 18 years on waiting lists with symptoms of depression, anxiety, or stress. Exclusion: Psychosis, use of other psychological service during treatment period.	149 73.8% female Age: M=35.3 (SD=10.3)	1. **iCBT** n=80 2. Wait-list control (**WL**) n=69	DASS-21 subscale 32 days	*Completer analysis only (no ITT)* <u>DASS-21 subscale at post-test (32 days)</u> iCBT vs WL: NSD (iCBT n=28, WL n=38)	Not reported

Table 3.2. Characteristics of studies included in the current systematic review (continued).

Study **Intervention (language)**	**Country** **Recruitment setting**	**Inclusion criteria** **Exclusion criteria**	**N** **% female** **Mean age (SD)**	**Trial conditions** [a]	**Depression measure(s)** **Assessment timepoints**	**Depression outcome iCBT vs control** [b] **Effect size (95% CI)**	**Intervention adherence**
Lillevoll KR, et al. 2014 [100] *MoodGYM* (Norwegian)	Norway Secondary school	Students at four senior high schools	775 56.8% female Age: M=16.8 (SD=1.00)	Due to error only 707 randomised: 1. Wait-list control (**WL**) n=180 2. iCBT no emails (**iCBT**) n=176 3. iCBT standard auto emails n=176 4. iCBT tailored auto emails n=175	CES-D 6 weeks	*Completer analysis only (irrespective of condition) (no ITT)* <u>CES-D at post-test (6 weeks)</u> Accessed iCBT (n=42) vs did not access (n=483): NSD *Sub-group with CES-D≥16:* Accessed iCBT (n=19) vs did not access (n=179): NSD	45 participants (8.54%) in an intervention conditions accessed the program.
Mohr DC, et al. 2013 [101] *moodManager* (English)	United States Primary care	Adults ≥ 18 years, MDD on MINI, QIDS ≥ 11 Exclusion: severe psychiatric disorder, dementia, substance or alcohol abuse, current psychological treatment or recent medication, suicidality	102 71.4% female Age: M=48.3 (SD=12.87)	1. iCBT with email and telephone coaching (GiCBT) n=34 2. **iCBT** n=35 3. Wait-list control (**WL**) n=33	PHQ-9 6 weeks 12 weeks˜	<u>PHQ-9 at post-test (6 weeks)</u> iCBT vs WL: d=0.41† (0.07, 0.89)†	Unguided iCBT group: Logged in mean 7.7 days, viewed mean of 5.6 lessons.

Table 3.2. Characteristics of studies included in the current systematic review (continued).

Study **Intervention (language)**	**Country** **Recruitment setting**	**Inclusion criteria** **Exclusion criteria**	**N** **% female** **Mean age (SD)**	**Trial conditions** [a]	**Depression measure(s)** **Assessment timepoints**	**Depression outcome iCBT vs control** [b] **Effect size (95% CI)**	**Intervention adherence**
Lintvedt OK, et al. 2013 [102] *MoodGYM* (Norwegian)	Norway University	University students (Adults ≥ 18 years), K-10 ≥ 20	163 76.7% female Age: M=28.2 (SD=7.4)	1. iCBT and BluePages psycho-education (**iCBT**) n=81 2. Wait-list control (**WL**) n=82	CES-D 8 weeks	<u>CES-D at post-test (8 weeks)</u> iCBT vs WL: d=0.57 (SE=0.16)	iCBT group: 47 participants (58.0%) accessed program, mean 2.8 modules.
Powell J, et al. 2012 [103] *MoodGYM* (English)	England Community (NHS Choices website)	Adults ≥ 18 years, living in England	3070 77.9% female Age: M=41.1 (SD=13.0)	1. **iCBT** n=1,534 2. Wait-list control (**WL**) n=1,536	CES-D 6 weeks 12 weeks	<u>CES-D at post-test (6 weeks)</u> iCBT vs WL: d=0.16† (0.09, 0.23)† <u>CES-D at 12 weeks</u> iCBT vs WL: d=0.19† (0.12, 0.26)†	*Based on graph:* iCBT group: approximately 60% ≥ 1 module; 33% ≥ 2 modules.
Berger T, et al. 2011 [104] *Deprexis* (German)	Switzerland and Germany Community (Internet and print advertising)	Adults ≥ 18 years, BDI-II > 13, MDD or dysthymia on MINI interview. Exclusion: suicidality, current treatment.	76 69.7% female Age: M=38.8 (SD=14)	1. Unguided iCBT (**iCBT**) n=25 2. Guided iCBT (GiCBT) n=25 3. Wait-list control (**WL**) n=26	BDI-II 10 weeks 6 months~	*Last observation carried forward* <u>BDI-II at post-test (10 weeks)</u> iCBT vs WL: d=0.66 (0.10, 1.23)† No effects of gender, age, education, previous treatment, medication, co-morbid mental health conditions. Dose-effect relationship not significant.	Unguided iCBT group: On average 6.8 lessons (SD=3.75) and spent 4hr 15 mins. 36% completed whole program.

Table 3.2. Characteristics of studies included in the current systematic review (continued).

Study **Intervention (language)**	**Country** **Recruitment setting**	**Inclusion criteria** **Exclusion criteria**	**N** **% female** **Mean age (SD)**	**Trial conditions** [a]	**Depression measure(s)** **Assessment timepoints**	**Depression outcome iCBT vs control** [b] **Effect size (95% CI)**	**Intervention adherence**
Farrer L, et al. 2011 [105] *MoodGYM* (English)	Australia National telephone helpline	Adults ≥ 18 years, K10 ≥ 22 Exclusion: suicidality, high distress, schizophrenia, bipolar disorder, current CBT treatment, significant reading impairment.	155 81.9% female Age: M=41.5 (SD=12.36)	1. Web Only (**iCBT**) n=38 2. Web + phone tracking n=45 3. Tracking only n=37 4. Treatment as usual (**TAU**) n=35	CES-D 6 weeks 6 months	<u>CES-D at post-test (6 weeks)</u> iCBT vs TAU: *g*=0.76 (0.21, 1.31) <u>CES-D at 6 months</u> iCBT vs TAU: *g*=1.19 (0.55, 1.83)	iCBT group: Mean no. modules: 1.5 (SD=1.89); 15.8% completed all modules; 31.6% completed ≥3 modules; 50.0% did not complete any modules.

Table 3.2. Characteristics of studies included in the current systematic review (continued).

Study **Intervention (language)**	**Country** **Recruitment setting**	**Inclusion criteria** **Exclusion criteria**	**N** **% female** **Mean age (SD)**	**Trial conditions** [a]	**Depression measure(s)** **Assessment timepoints**	**Depression outcome iCBT vs control** [b] **Effect size (95% CI)**	**Intervention adherence**
Meyer B, et al. 2009 [106] *Deprexis* (German)	Germany Community (Internet advertising, including depression forum notices)	Adults ≥ 18 years	396 76% female Age: M=34.8 (SD=11.60)	1. Immediate treatment (**iCBT**) n=320 2. Delayed treatment (**WL**) n=76	BDI 9 weeks 18 weeks˜ 6 months ˜	*Last observation carried forward* BDI at post-test (9 weeks) iCBT vs WL: d=0.30 (0.05, 0.55)	Of 396 participants (both conditions had iCBT access after 9 weeks): 78.3% ≥ 1 session (>10mins); 62.9% 2 sessions; 46.2% 3 sessions.
de Graaf LE, et al. 2009 [107] *Colour Your Life* (Dutch)	Netherlands Community (invitation to random selection of general population)	Adults 18-65 years, BDI-II ≥ 16, depressive symptoms > 3 months. Exclusion: current treatment, medication, alcohol/ drug dependence; severe psychiatric comorbidity.	303 56.8% female Age: M=44.9 (SD=11.62)	1. **iCBT** n=100 2. TAU by GP (**TAU**) n=103 3. iCBT and TAU combined n=100	BDI-II 2 months 3 months 6 months	*Intermittent data imputed; missing data due to loss to follow-up not imputed.* BDI-II at post-test (2 months) iCBT vs TAU: NSD BDI-II at 3 months iCBT vs TAU: NSD BDI-II at 6 months iCBT vs TAU: NSD	iCBT group: 72.0% completed first session 14.0% completed all sessions Mean no. sessions: 3.4 (SD=3.0).

Table 3.2. Characteristics of studies included in the current systematic review (continued).

Study **Intervention (language)**	**Country** **Recruitment setting**	**Inclusion criteria** **Exclusion criteria**	**N** **% female** **Mean age (SD)**	**Trial conditions** [a]	**Depression measure(s)** **Assessment timepoints**	**Depression outcome iCBT vs control** [b] **Effect size (95% CI)**	**Intervention adherence**
Spek V, et al. 2007 [108, 109] iCBT (Dutch) 8 modules based on Coping With Depression	Netherlands Community (print advertisements and letters)	Adults 50-75 years with EDS ≥ 12 but not diagnosis of depression. Exclusion: other psychiatric disorder in need of treatment, suicidal ideation.	301 63.9% female Age: M=55 (SD=4.6)	1. **iCBT** n=102 2. Group intervention n=99 3. Wait-list control (**WL**) n=100	BDI-II 10 weeks 12 months	*Effect sizes calculated as difference between within-group effect sizes* <u>BDI-II at 10 weeks</u> iCBT vs WL: d=0.55 (significant, CI/p not reported) <u>BDI-II at 12 months</u> iCBT vs WL: d=0.53 (significant, CI/p not reported)	iCBT group: Mean modules completed (of 8): 5.5. 48.3% completed all modules. Dose-effect relationship: NS
Clarke G, et al. 2002 [16] *ODIN* (English)	United States Secondary care (Health maintenance organisation)	Adults ≥18 years (received depression treatment in last 30 days or depression diagnosis; or from non-depressed gender and age-matched group)	299 75.6% female Age: M=43.9 (SD=12.30)	1. **iCBT** n=144 2. Treatment as usual (**TAU**) n=155	CES-D 4 weeks 8 weeks 16 weeks 32 weeks.	CES-D at follow-up time-points iCBT vs TAU: NSD. Participants with CES-D≥ 20: NSD *Participants with CES-D<20:* At 16 weeks: d=0.17 (p<.05) At 32 weeks: d=0.48 (p<.01).	iCBT group: Mean no. sessions: 2.6 (SD=3.5). Dose-effect relationship: NS

[a] Control condition and automated iCBT condition bolded.

[b] Significant between-group effect size (based on intention to treat analysis) unless otherwise specified. Cohen's d calculated using post-test estimated means and pooled SD at post-test unless otherwise specified.

˜All participants had received access to the intervention at this time-point.

† Calculated by the current author based on reported data.

CI= Confidence interval; iCBT= Internet-delivered cognitive behaviour therapy; WL = Wait-list control group; TAU = Treatment as usual; NSD= No significant difference; NS= Not significant; BDI= Beck Depression Inventory; CES-D= Center for Epidemiological Studies Depression Scale; PHQ-9= Patient Health Questionnaire (9 items); IDS-SR= Inventory of Depressive Symptomatology Self-Report; DASS-21= Depression, Anxiety and Stress Scale (21 items); QIDS= Quick Inventory of Depressive Symptomology; K10= Kessler Psychological Distress Scale; MINI: Mini Diagnostic Interview for Psychiatric Disorder.

Year of publication

The earliest trial of automated iCBT was published in 2002 [16], and the majority of studies were first published in 2013 or later (*n* = 10, 58.8%). The year of publication for included studies (of the first paper where more than one has been published for a trial) is shown in Figure 3.2.

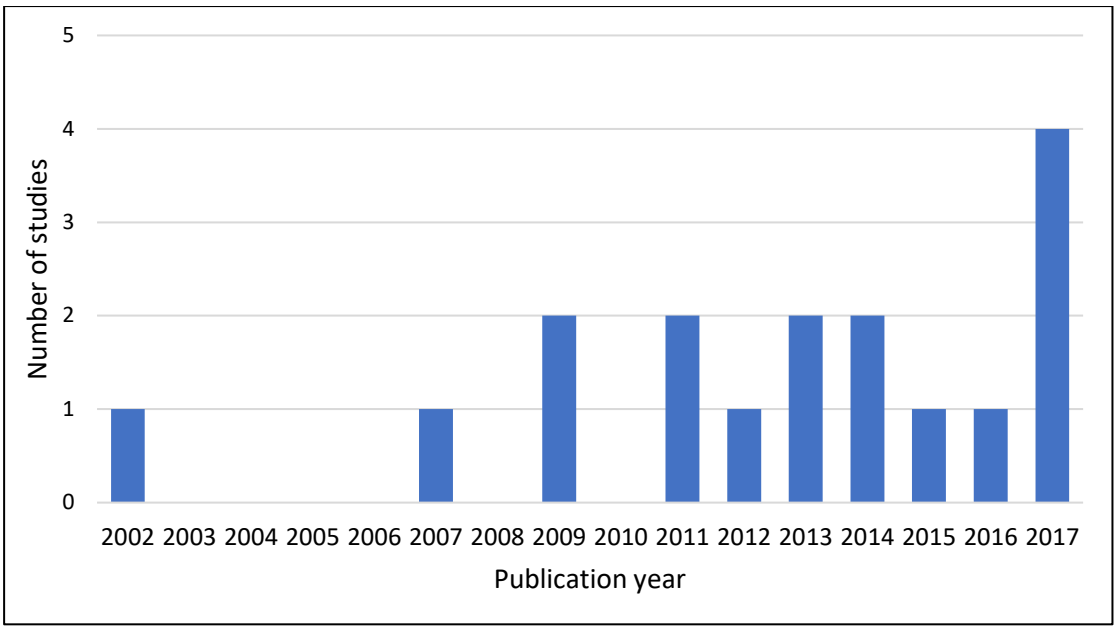

Figure 3.2. Year of publication (first article) for studies included in the systematic review.

Location

The largest number of studies were conducted in the Netherlands (*n* = 3) [94, 107, 108], followed by two studies in each of Germany [98, 106], Spain [92, 97], the United States [16, 101] and Norway [100, 102]. One study recruited participants residing in both Germany and Switzerland [104], and the remaining studies were conducted in Japan [93], Australia [105], Canada [95], Ireland [99] and England [103].

Settings

Over half of the studies (*n* = 9, 52.9%) recruited participants from the community [92, 93, 94, 98, 103, 104, 106, 107, 108]. Nearly all of these (*n* = 8) employed Internet- or print-based advertisements that encouraged interested individuals to read more about the study and to progress to the screening phase of the study, and one recruited participants from the community using random sampling and individual invitations [107]. One trial, which has been classified as 'community-based' in the current review, recruited from advertisements in in-patient and out-patient clinics in addition to advertisements in newspapers and Internet forums [98]. Three trials directly recruited primary care patients [95, 97, 101], and two trials recruited secondary care patients (one of these studies involved

individuals on waiting lists for mental health services [99], and the other engaged members of a health maintenance organisation in the United States [16]). One trial recruited participants from callers to a national crisis telephone helpline [105]. The remaining two trials were conducted with senior secondary school students [100] and with university students [102].

Study populations

Demographic status: All but one of the RCTs involved adult participants aged 18 years or over (*n* = 16, 94.1%). Of the trials involving adults, one targeted older adults aged 50-75 years [108] and one recruited university students with a mean age of 28.2 (*SD* = 7.4) [102]. The remaining trial involved senior secondary school students, with a mean age of 16.8 years (*SD* = 1.0) [100]. The percentage of female participants ranged from 50.1% to 81.9%. There was a predominance of women in most trials, with the sample comprising more than 60% women in 13 of the 16 trials where data on gender were available.

Inclusion criteria: The inclusion criteria for half of the trials (*n* = 9, 52.9%) required participants to have either a depression diagnosis through a phone-administered Mini-International Neuropsychiatric Interview (M.I.N.I.) [110] (*n* = 4) [92, 97, 101, 104] or to have clinically significant depression symptoms on a validated depression scale (*n* = 5) [93, 94, 102, 105, 107]. The latter included two studies with inclusion criteria based on the Kessler Psychological Distress Scale (K10) [102, 105] since the K10 has been validated as indicating either a depressive or an anxiety condition [111].

Three additional studies identified participants with depression or who were at risk of depression using less rigorous inclusion criteria for assessing depression as follows. Two trials targeted participants at risk of depression but did not impose a requirement that participants score higher than a minimum cut-off on a validated scale. One of the latter studies included participants experiencing a current stressful event interfering with daily functioning (but who scored less than 28 on the Beck Depression Inventory-II) [92] and one recruited from individuals on a waiting list who were referred to public mental health services for depression, anxiety or stress issues [99]. A third RCT recruited participants who had recently received depression treatment or had a depression diagnosis provided by a medical practitioner, along with a gender- and age-matched group of non-depressed participants [16].

Finally, five studies could be categorised as prevention trials [95, 100, 103, 106, 108] and involved inclusion criteria as follows. One trial specifically targeted participants with sub-threshold depression symptoms only, excluding individuals who met criteria for depression diagnosis determined through

a telephone interview (indicated preventive intervention) [108]. Four RCTs did not impose any restriction on the depression status of included participants (universal intervention) [95, 100, 103, 106].

Interventions

Six RCTs investigated the use of the 'MoodGYM' intervention that provides a 5-module CBT training program [95, 99, 100, 102, 103, 105]. Four of these studies employed the English version of *MoodGYM* [95, 99, 103, 105], and two were undertaken using the Norwegian version [100, 102]. The next most frequently studied intervention (*n* = 3) was 'Deprexis', a German language CBT intervention comprising ten modules [98, 104, 106]. A Spanish intervention 'Smiling is Fun', which is organised into ten modules, was studied in two trials [92, 97]. The remaining studies each investigated a different iCBT intervention. These were: 'ODIN' (Overcoming Depression on the Internet), an English language program divided into 7 chapters [16]; 'moodManager', an English language program divided into 18 short lessons [101]; and 'Colour Your Life', a Dutch program comprising eight 30-minute sessions plus boosters [107]. Three interventions were unnamed. These included an eight-module Dutch program based on the 'Coping With Depression' CBT course [108]; a suite of three Complaint Directed Mini-interventions based on CBT (Dutch language) [94]; and a five minute simplified iCBT intervention (Japanese language) [93].

Control conditions

The majority of the 17 studies (*n* = 11, 64.7%) compared the outcomes of participants in the automated iCBT intervention group to those in a wait-list control group [92, 93, 94, 99, 100, 101, 102, 103, 104, 106, 108]. In five RCTs, participants in the control condition received treatment as usual [16, 95, 98, 105, 107], and in an additional trial the control participants received 'improved' treatment as usual from a GP who had received three hours training in the treatment of depression as part of the trial protocol [97].

Human involvement in inclusion/exclusion criteria screening

Five trials (29.4%) used human-based assessment of mental health symptoms (i.e., at screening via M.I.N.I.) to determine if individuals were eligible to participate in the study [92, 97, 98, 101, 104]. Mental health symptom screening in all other trials (76.5%) was conducted online (automated, self-report questionnaires).

Depression outcome measures

In all trials, depression outcomes were measured via self-report at baseline and at post-intervention time-points. Depression severity was most commonly measured using the Beck Depression Inventory (BDI, 21 items; *n* = 6, 35.3%) [92, 97, 104, 106, 107, 108], with all but one of these trials using the BDI-II (1996 revision of the scale) [106]. The 20 item Center for Epidemiological Studies Depression Scale (CES-D) was also used in six trials (35.3%), and the nine-question Patient Health Questionnaire (PHQ-9) in three trials [95, 98, 101]. One trial used the depression sub-scale of the Depression Anxiety Stress Scales (DASS-21) [99] and one trial used the Inventory of Depressive Symptomatology Self-Report (IDS-SR) [94].

3.3.3 Quality of studies

Overall the risk of bias in the 17 included trials was low, with 13 studies (76.5%) being assessed as adequate against all four quality criteria (see Table 3.3). All studies used self-report measures for outcomes, and so met the quality criteria for blinding of outcome assessors. Two studies failed to describe an adequate procedure for generation of the randomisation sequence, and three studies did not report that the allocation of condition to participants was concealed (for example, that it was undertaken by a third party). In addition, three studies reported completer analyses but not intent to treat findings, introducing the risk of attrition bias.

Table 3.3. Assessment of quality criteria for studies included in the current systematic review.

	Allocation sequence generation	Concealment of allocation (3rd party)	Blinding of outcome assessors	ITT analysis
Mira A, et al. 2017 [92]	+	+	+	+
Noguchi R, et al. 2017 [93]	+	+	+	+
Lokman S, et al. 2017 [94]	+	+	+	+
Silverstone P, et al. 2017 [95]	-	-	+	-
Montero-Marin J, et al. 2016 [97]	+	+	+	+
Meyer B, et al. 2015 [98]	+	+	+	+
Twomey C, et al. 2014 [99]	+	-	+	-
Lillevoll KR, et al. 2014 [100]	+	-	+	-
Mohr DC, et al. 2013 [101]	+	+	+	+
Lintvedt OK, et al. 2013 [102]	+	+	+	+
Powell J, et al. 2012 [103]	+	+	+	+
Berger T, et al. 2011 [104]	+	+	+	+
Farrer L, et al. 2011 [105]	+	+	+	+
Meyer B, et al. 2009 [106]	+	+	+	+
de Graaf LE, et al. 2009 [107]	-	+	+	+
Spek V, et al. 2007 [108]	+	+	+	+
Clarke G, et al. 2002 [16]	+	+	+	+

3.3.4 Effectiveness of automated iCBT for depression

The effectiveness of the automated iCBT conditions compared to control conditions in reducing depressive symptoms are summarised in Table 3.2, along with details of participant adherence to the iCBT intervention. A significant intervention effect was detected at post-intervention in 11 of the 17 included trials (64.7%), with effect sizes ranging from 0.16 to 0.76. Depression outcomes at later time-points (up to 12 months) for automated iCBT conditions compared to control conditions were also reported in seven trials, with significant intervention effects in five (71.4%). The details and outcomes of each trial are described below, organised by the intervention used in the study.

Studies of MoodGYM (English and Norwegian)

Six trials studied the effectiveness of the *MoodGYM* training program against a control condition, four using the English version [95, 99, 103, 105] and two the Norwegian version [100, 102]. Of these, three trials – one involving telephone helpline callers, one targeting university students, and one comprising community-based visitors to an online government health portal – yielded positive outcomes [102, 103, 105].

In the first of these, Farrer and colleagues investigated the effectiveness of the English-language program for adult callers to an Australian telephone helpline who had a score of 22 or above on the Kessler Psychological Distress Scale (K10) [105]. The included participants (*n* = 155) were randomised to one of four conditions, including one that was an automated iCBT condition and another that was a treatment as usual control condition. Analysis was conducted on an ITT basis using mixed modelling. The automated iCBT group showed significantly greater improvement on the CES-D compared to the control group at 6-week post-test (Hedge's *g* = 0.76, 95% CI: 0.21, 1.31) and there was a large treatment effect after 6 months (*g* = 1.19, 95% CI: 0.55, 1.83). On average, participants in the iCBT condition completed 1.5 modules (*SD* = 1.89). Nearly one-third of these participants (31.6%) completed three or more modules and 15.8% completed all five modules; however, 50.0% did not complete the first module.

Lintvedt and colleagues also studied the use of *MoodGYM* in a population of adults with elevated K10 [102]. In this study of the Norwegian version of *MoodGYM*, potential participants were identified by sending letters and screening surveys to all students at a Norwegian university. Those with K10 scores of 20 or above were eligible for inclusion in the study. A total of 163 participants were randomised into an automated iCBT group or a wait-list control. The results of a mixed model analysis showed a significant reduction of depressive symptoms (CES-D) in the intervention group, with a medium effect size at post-test (*d* = 0.57, *SE* = 0.16). More than half of the iCBT group (58.0%) accessed *MoodGYM* (on average, 2.8 modules).

The English version of *MoodGYM* was also studied in a large community sample with no restriction on depression levels [103]. In this trial, adults residing in England were recruited to a study about wellbeing via an advertisement on the NHS Choices website in England. The primary outcome for the trial was mental well-being (measured on the Warwick-Edinburgh Mental Well-being Scale) and depression severity (CES-D) was a secondary outcome measure. Participants (*n* = 3,070) were randomised to receive either *MoodGYM* (with weekly automated reminder emails) or a wait-list control and were invited to complete post-test surveys after 6 and 12 weeks. At each of these time-points, an ITT analysis showed there was a significant reduction in depression severity for

participants in the intervention group compared to the control. Calculated using the difference in estimated marginal means and the standard errors at the later time-point, the authors reported a Cohen's *d* effect size of 0.17 (95% CI: 0.10, 0.24) at 6 weeks and 0.19 (95% CI: 0.12, 0.26) at 12 weeks. Based on a graph in the published article, approximately 60% of the intervention group completed at least one module, and around one third completed two modules. There were no analyses of the predictors of depression improvement; however, there were significant improvements in wellbeing for participants who completed two or more modules.

Three trials of automated *MoodGYM* failed to find evidence of effectiveness of the program [95, 99, 100]. Two of these trials involved participants recruited from clinical settings; one involved a universal sample (no restriction on depression status) of students recruited from senior high schools.

In the first of these, Twomey and colleagues investigated the use of *MoodGYM* with clients on a waiting list for public mental health services in Ireland [99]. Participants were 149 adults who had been referred to face-to-face services for depression, anxiety or stress, randomised to receive either *MoodGYM* or to a wait-list control condition in which participants were offered access to *MoodGYM* after 12 weeks. Depression severity was measured using the depression sub-scale of the DASS-21 with post-test at 32 days. The analysis of differential effectiveness was conducted on a completer basis (including data from only 66 participants) and did not show a significant difference between the intervention and wait-list groups in the reduction of depression. No adherence data were reported.

In a trial conducted in primary care in Canada, Silverstone et al. recruited 1,489 medical practice patients each of whom subsequently completed the PHQ-9 questionnaire [95] and was assigned to one of four assigned conditions. Of these conditions, one was 'treatment as usual (TAU)', and one was 'treatment as usual plus access to *MoodGYM* (iCBT)'. Randomisation was conducted at the level of an entire day of patients at a particular clinic so that all potential participants in the clinic waiting room were provided with the same instructions. Participants were followed up after 12 weeks, and a completer analysis was conducted. No significant difference in depression severity on the PHQ-9 was detected between the iCBT and TAU conditions in the completer sample at post-test (*n* = 530). Nor was there a difference when the analysis was confined to participants with a baseline PHQ-9 score of 10 or more (clinically depressed cut-off score) (*n* = 72). Of the 29 participants with high baseline PHQ-9 scores who completed the post-test, 25 accessed the program at least once, but fewer than five completed more than one session.

Finally, a study of students recruited from senior high schools in Norway failed to find evidence of effectiveness of the Norwegian version of *MoodGYM*. In this study, students were eligible for

inclusion in the study regardless of their depression status. Approximately half (52.8%) of the students who were offered participation in the trial consented to participate, and 707 participants were randomised to one of four conditions: *MoodGYM* (iCBT), *MoodGYM* plus a standard weekly email reminder, *MoodGYM* plus a tailored weekly email reminder and a wait-list control. All three of the intervention groups comprised automated iCBT since both email types were automated, and there was no human contact with participants. Adherence to the intervention in this trial was unexpectedly low, with only 45 of the 527 (8.54%) participants in the intervention groups accessing the program at least once. Analysis of the effectiveness of the intervention was conducted on a completer basis, comparing outcomes for those participants who accessed the program (any intervention group) and those who did not (any condition), for the pool of participants who completed the post-test. No significant effect of *MoodGYM* use was found, even when the analysis was restricted to participants with a baseline CES-D score of 16 or more.

Studies of Deprexis (German)

Three studies used the German *Deprexis* iCBT intervention as part of a purely automated condition, and all three yielded positive effects at post-test [98, 104, 106]. Two of the trials employed community samples [104, 106] and the third recruited from both the community and clinics [98].

In a study of automated *Deprexis*, published in 2009, Meyer and colleagues recruited adults from the community (through advertisements on depression Internet forums) with no restriction on symptoms (n = 396) and randomised participants to either the iCBT condition or a wait-list condition which received the intervention 9 weeks later [106]. The ITT analysis of the effectiveness of the intervention on depression (BDI score) was conducted using a last observation carried forward model which resulted in an effect size of d = 0.30 (95% CI: 0.05, 0.55). Of the total number of participants (including those in the wait-list condition), 78.3% accessed the intervention at least once for more than 10 minutes, 62.9% at least twice, and 46.2% on at least three occasions.

A subsequent trial investigated the use of *Deprexis* among community-recruited adults with either Major Depressive Disorder or Dysthymia, as determined through a telephone-delivered interview [104]. In this trial, 76 participants were randomised to either automated iCBT, a therapist guided iCBT condition or a wait-list control. At post-test (10 weeks) the Cohen's d effect size between the automated iCBT group and the control group was 0.66 (95% CI: 0.10, 1.23). This analysis handled missing post-test scores using the last observation carried forward method. Those in the unguided iCBT condition completed an average of 6.8 of 10 lessons (SD = 3.75), and 36% completed the whole program. Potential baseline predictors of depression improvement were investigated, with no

significant relationships detected. The number of lessons completed did not significantly correlate with improved depression outcomes.

The final trial of *Deprexis* involved participants with severe depression symptoms (a score of more than 14 on the PHQ-9) and who were diagnosed with either Major Depressive Disorder or Dysthymia on the basis of a telephone interview [98]. Participants in this trial were recruited from those excluded from a larger, parent trial of *Deprexis* due to the severity of their depression symptoms. (The parent trial was not eligible for inclusion in the current review because it incorporated therapist guidance for participants with a PHQ-9 score between 10 and 14 [112, 113].) For the automated iCBT trial, 78 participants were randomised to receive access to *Deprexis*, and 85 participants continued with treatment as usual. Post-test measures were administered via an online survey at 3 and 6 months. At 3 months, a medium between-groups effect size was reported (d = 0.57, 95% CI: 0.22, 0.92); however, the treatment effect was not significant at 6 months (d = 0.33, 95% CI: -0.03, 0.69). These effect sizes were based on *observed* means, with no imputation. The authors defined treatment adherence as engagement with the intervention in at least four sessions for a minimum of 60 minutes, criteria which were met by 56 (71.5%) of participants in the iCBT condition. Predictor analysis using a linear mixed effects model identified a significant three-way interaction between anti-depressant use and group and time. Specifically, participants in the iCBT group who were taking anti-depressants demonstrated significantly greater improvement in depression symptoms on the PHQ-9 than those who were not whereas no such effect was demonstrated for the TAU group. There were no other significant interactions between baseline variables and group and time.

Studies of Smiling is Fun (Spanish)

Two studies, one in a primary care setting and one in a community setting, investigated the effectiveness of the automated *Smiling as Fun* iCBT intervention [92, 97]. Each yielded positive outcomes at one or more time-points.

The first of these was published in 2016 and reported on the use of the intervention in a primary care sample in Spain [97]. Participants were adult patients with mild to moderate depression symptoms (measured on the BDI-II), who were diagnosed with Major Depressive Disorder through a M.I.N.I. telephone interview and recruited from 46 GP clinics across Spain. There were three trial conditions of which one was 'improved' treatment by a GP who had undertaken three hours specialist training in the diagnosis and treatment of depression (control condition), and one was improved treatment by a GP plus unguided access to the automated iCBT intervention. An ITT multi-level mixed effects model for repeated measures was constructed with data from all 296 participants, adjusted for baseline BDI-II scores, with regression coefficients reported. There was no significant difference in

BDI-II scores at 3 months for the iCBT group compared to the control group; however, there was a significant difference in favour of iCBT at 6 months (B = 2.91, p < .001) and at 12 months (B = 3.69, p < .001). Effect sizes based on observed data (n = 188) were reported, with a very small and not significant iCBT intervention effect at 3 months (Hedge's g = 0.12), and significant small to medium effect sizes at 6 months (Hedge's g = 0.34, p = .007) and at 12 months (g = 0.48, p = .001). After 12 months, 72.4% of the iCBT group had accessed the intervention, and 41.8% had completed more than six of the ten modules. An economic evaluation from a societal perspective found that iCBT was likely to be more effective and less expensive than improved treatment by a GP [96]. Cost-effectiveness analyses, using an ITT approach, resulted in an incremental cost-effectiveness ratio (ICER) of -€98.37 for iCBT compared to improved treatment as usual, meaning that each point of improvement on the BDI-II achieved using the iCBT condition saved €98.37 compared to using TAU. The corresponding cost-utility analysis reported an incremental cost-utility ratio (ICUR) of €5,160.40, representing the saving achieved with each extra quality-adjusted life year (QALY) gained using iCBT compared to improved treatment by a GP.

The second *Smiling is Fun* intervention trial was undertaken on a self-selected community sample (N = 132) of individuals experiencing at least one stressful event which caused interference in daily life [92]. Participants were randomly assigned to either a wait-list control or to one of two interventions, one being an automated iCBT. A medium effect size was reported between the automated iCBT and wait-list control groups at 12-week post-test on the BDI-II (d = 0.50, 95% CI: 0.05, 0.94). All participants received access to the iCBT intervention after 12 weeks, precluding a meaningful 12-month follow-up comparison between groups. However, the within-group effect size for the automated iCBT group remained medium at 12 months (d = 0.45, 95% CI: 0.13, 0.76). Over three-quarters of participants in this group completed all modules. Participants in the automated iCBT condition completed, on average, 77.8% of the program, and there was no significant difference in the intervention adherence rate for the two intervention conditions.

Studies of other iCBT interventions (English, Dutch and Japanese)

The remaining six studies investigated the effectiveness of different iCBT interventions. They included three Dutch studies which recruited participants from the community [94, 107, 108], two studies of English-language interventions in primary and secondary care settings [16, 101], and one community-based Japanese study that used a brief 5-minute iCBT intervention [93]. Four trials reported significant intervention effects at post-intervention [93, 94, 101, 108], and in one trial these effects were sustained at 12 months [108]. No significant between-group effects for depression were

reported for the remaining two studies [16, 107]. Here we consider the studies yielding positive outcomes first followed by a description of the negative trials.

The single study of the *moodManager* iCBT program, published in 2013, yielded a positive outcome [101]. The trial comprised 102 adults with depression from primary care with the depression diagnosis confirmed using the M.I.N.I. Participants were randomised to either a wait-list control group, an automated iCBT condition or another intervention condition. Calculated using reported results from a mixed model analysis, at 6-week post-test there was an effect size between the automated iCBT group and the wait-list control group of 0.41 (95% CI: 0.07, 0.89) on the PHQ-9. Those in the automated iCBT condition, on average, accessed the program on 7.7 days and viewed a mean of 5.6 (of a total of 18) lessons. After controlling for baseline PHQ-9 score and group allocation, increased use of the program was reported to significantly predict improvement in depression symptoms after 12 weeks. At this time-point, all participants had received access to the iCBT program, with control condition participants being provided with a choice of either automated or guided iCBT after 6 weeks.

Of the remaining trials which yielded positive outcomes, one 2017 Dutch study investigated the effectiveness of brief, 'complaint-directed mini-interventions' for reducing depressive symptoms [94]. The intervention consisted of online programs comprising three to four short modules, each targeting a different problem associated with depression (sleep, stress and worry). Participants were recruited from the community via advertisements and screened for depression. Those with mild to moderate depression (n = 329) were randomised to receive either access to the interventions or to a wait-list control. Participants in the intervention group could choose their program, with the largest number choosing the worry program (37.0%), followed by the sleep program (35.8%). At 3-month post-test, linear mixed model analysis showed a significantly greater reduction in depression symptoms (measured by the IDS-SR) in the intervention condition compared to the control condition. The paper reported a between-group Cohen's d effect size of 0.70 (95% CI: 0.38, 1.03) in an adjusted model controlling for age, gender, living arrangement, education and baseline depression and anxiety scores. This effect size was computed using the estimated mean *differences* in depression scores and the pooled standard deviation at baseline. However, when calculated by the current author, using the post-test estimated means and standard error, Cohen's d was 0.39 (95% CI: 0.17, 0.61).

Noguchi and colleagues also reported some evidence of a short-term positive effect of a very brief intervention for adults with depression. The intervention comprised a five-minute simplified iCBT program. Participants were 974 Japanese adults with at least mild depression (scores of 16 or more

on the CES-D and 9 or more on the PHQ-9) [93]. Participants were recruited from a marketing company website, and the sample included the largest proportion of males (49.9%) of the trials included in the current review. Participants were randomly allocated to either a wait-list control, the iCBT program or another intervention. Daily automated emails were sent to the intervention groups, and participants in these groups were encouraged to access their allocated program at least twice per week for 5 weeks. A linear mixed model analysis showed no significant difference between the iCBT and the control group in improvement on the CES-D or the PHQ-9 at 6-week post-test or at 12-week follow-up. However, there was a time delay between screening and baseline assessment, such that 195 of the participants who were randomised no longer met the minimum depression severity inclusion criteria at baseline. When these participants were excluded, the outcome analysis revealed a significant iCBT intervention effect at 6 weeks on both the CES-D and PHQ-9, but there was no significant difference at 12 weeks. Based on the reported estimated means from this analysis, the between-group effect size at 6-week post-test was calculated by the current author for both the CES-D (d = 0.34, 95% CI: 0.16, 0.51) and the PHQ-9 (d = 0.27, 95% CI: 0.10, 0.44). No data were provided on the frequency with which participants accessed the brief intervention.

Finally, an indicated prevention trial yielded positive outcomes at post-test and long-term follow-up. This Dutch study, published in 2007, focused on adults aged 50 to 75 years with sub-threshold depression recruited from the community [108, 109]. In this study, 301 participants were randomised to receive an eight-module iCBT intervention, another CBT intervention or a waiting list control. At post-test (10 weeks), there was a significantly greater improvement in depression among the iCBT group than the control group (d = 0.55, CI not reported). At 12-month follow-up, there was also a significant intervention effect for iCBT compared to control (d = 0.53, CI not reported). These effect sizes were calculated by subtracting the within-group effect size for the control group from that of the intervention group. Nearly half of the participants in the iCBT condition completed all modules (48.3%), with participants completing a mean of 5.5 modules.

Of the two negative trials, one investigated the effectiveness of *Overcoming Depression on the Internet* (*ODIN*) program. This trial – which was the earliest randomised control trial of an automated iCBT – was conducted in 2002, with 299 adults who were recruited from a health maintenance organisation in the U.S. [16]. Potential participants were chosen either because they had recently received depression treatment or had a depression diagnosis or were not depressed but were part of a group selected to match the age and gender characteristics of the depressed group. Participants were followed up using online surveys at 4, 8, 16 and 32 weeks. There was no significant difference in the CES-D scores for participants in the iCBT group compared to the control group at any time-point.

When the group was separated into those with high baseline depression scores (CES-D ≥ 20) and low baseline scores (CES-D < 20), no significant intervention effect was detected in the highly depressed group (*n* = 236). However, there was a significant treatment effect at 16 weeks (*d* = 0.17, *p* < .05) and at 32 weeks (*d* = 0.48, *p* <.01) in those with low baseline depression (*n* = 63). Participants in the iCBT group accessed the intervention on average 2.6 times (*SD* = 3.5); however, there was no significant effect of the number of sessions on depression improvement.

Finally, a 2009 study of the Dutch Internet depression intervention *Colour Your Life* failed to find evidence of effectiveness [107]. A random community sample was screened for depression with those with a score of 16 or more on the BDI-II and depression symptoms for more than 3 months eligible for inclusion in the trial. A total of 303 participants were randomised to receive either depression treatment as usual by a GP (control), or one of two interventions, one being the automated *Colour Your Life* (iCBT) intervention. Outcome analysis for this study was reported as being on an intention to treat basis, but not all missing data were imputed. There was no significant time x group interaction and only very small differences between the within-group effect sizes of the iCBT and control conditions were reported (0.08 at 2 months, 0.02 at 3 months, 0.05 at 6 months). Effect sizes calculated by the current study author (based on the difference in the mean post-test scores divided by the pooled standard deviation at post-test) were small and non-significant at all time-points (2 months: *d* = 0.15, 95% CI: -0.13, 0.42; 3 months: *d* = 0.09, 95% CI: -0.19, 0.37; 6 months: *d* = 0.10, 95% CI: -0.18, 0.37). In this trial, only 14.0% of participants in the automated intervention condition completed all of the iCBT program.

3.4 Discussion

3.4.1 Summary of findings

The majority of studies included in this systematic review (*n* = 11, 64.7%) reported that an automated iCBT intervention was effective in the reduction of depression symptoms relative to a control condition at post-test assessment. Effect sizes for these positive trials ranged from 0.16 to 0.76. All of the trials reporting positive results were conducted with adult participants, the majority of which (*n* = 8) recruited participants from the community, with the remaining three trials targeting university students (*n* = 1), telephone helpline callers (*n* = 1) and primary care patients (*n* = 1). Seven of these positive trials included only participants with clinical depression symptom levels (determined through either a formal diagnostic telephone interview or scores on a validated scale), one included participants who were exposed to a stressful event that was interfering with their everyday life, one

included only older participants with sub-threshold depression symptoms and two had no restriction on the depression status of included participants.

The level of participant adherence to the automated iCBT interventions varied across studies. Studies differed with respect to the type of adherence data reported precluding a comprehensive analysis of the data across all studies. However, of the seven studies that reported how much of the program participants completed on average, adherence ranged from 30% in the trial of *MoodGYM* in a national telephone helpline setting to 68.8% in a community-based trial of *Deprexis*. Of the six studies that reported the number of participants who completed the entire intervention program, completion rates ranged from 0% in the Canadian primary care trial of *MoodGYM* to 77.8% in a trial of *Smiling is Fun* in a Spanish community setting. The mean number of logins ranged from 2.6 to 7.7 in the three trials that reported this information. The percentage of participants who did not once access the iCBT intervention was reported in six trials and ranged from 13.8% of depressed participants in the primary care trial of *MoodGYM* in Canada to 91.5% of secondary school students participating in a Norwegian trial of the same program. Two studies did not report any adherence data.

3.4.2 The effectiveness of self-selected automated iCBT in community settings

The results provide convincing evidence of the effectiveness of automated iCBT interventions for depression among self-selected individuals from the community. Of the nine community-based trials, the only study that did not demonstrate effectiveness involved a randomly selected community sample screened for depression [107]. The remaining eight community trials used recruitment methods such as online and print advertising that required participants to explicitly register their interest in these studies and to self-select participation [92, 93, 94, 98, 103, 104, 106, 108]. Thus, these individuals may have been more motivated and engaged with the intervention than a randomly selected, individually invited sample.

Of the four self-selected community trials (n = 4) that targeted participants with a depression diagnosis or clinically significant symptoms of depression (treatment trials), two established depression status via a human-conducted structured diagnostic telephone interview [98, 104] and two employed automated screening [93, 94]. The magnitudes of the effect sizes at post-test for the two treatment trials that employed telephone assessment during eligibility screening were somewhat larger than those for the two trials that used automated screening (telephone screening: d = 0.57 [98], d = 0.66 [104]; automated screening d = 0.39 [94]; d = 0.34 [93]). Such differences may be attributable to a range of factors such as the nature of the intervention and the setting (both trials

involving telephone screening used the *Deprexis* program in users in Germany, whereas those using automated screening involved a Dutch complaint-directed intervention suite and a brief Japanese intervention advertised on a marketing website). Further, the trials each used a different depression scale as the outcome measure. However, it is also possible that human contact at screening encouraged participation and subsequent engagement with the intervention content. Adherence to the programs is difficult to compare between the trials due to the lack of consistency in the variables that are reported in different trials. The participants in one *Deprexis* trial completed, on average 68% of the modules, and in the other 71.8% completed at least four sessions. In the Dutch trial, users logged on a median of 3 times and the Japanese study did not report adherence data. Future research employing head-to-head trials of the two recruitment methods may shed further light on this issue.

The largest community trial of automated iCBT, undertaken by Powell and colleagues, and with no restriction on the symptomology of participants, clearly demonstrated the potential public health benefit of automated iCBT [103]. Participants were recruited quickly and in large numbers (*n* = 3,070) through a major online government health website in response to a banner advertisement about a trial to increase mental well-being. Even without specifically targeting individuals with depression symptoms, there were small, but significant reductions in depression at 6 and 12 weeks relative to control. Given the scalability of fully automated iCBT, even small effect sizes could translate into significant public health gains by reducing depression symptoms in a large number of individuals.

3.4.3 The effectiveness of iCBT in primary care

The potential for the use of automated iCBT for the treatment of patients with depressive symptoms in primary care is less promising. Few participants used the iCBT intervention in the trial conducted in primary care in Canada (and there was no significant intervention effect) [95], and in the Spanish primary care trial there was no significant effect for iCBT compared to improved treatment as usual by a GP at 3-month post-test [97]. Both trials recruited participants during their visit to the primary care doctor – either in the waiting room (Canadian trial) or during the GP consultation (Spanish trial). It is possible that individuals with depressive symptoms who present to primary care are seeking face-to-face or guided treatment from their physician and are not motivated to undertake a self-directed, automated intervention. It is also possible there is minimal room for improvement due to the adjunct use of iCBT when the TAU treatments include evidence-based treatments such as anti-depressant medication. For example, in the Spanish trial, the majority of participants were prescribed medication (control group: 80.6%; iCBT group: 68.3%) with no significant difference between the prescribing rates for the groups. The Canadian trial did not report treatment use.

Interestingly, however, the Spanish trial did report a significant improvement in depression scores among the iCBT group compared to the control group at later time-points (after 6 and 12 months). This finding is consistent with evidence that CBT may be more effective over the long-term [114].

The third primary care trial did result in a significant effect size of 0.41 for the automated iCBT group compared to a wait-list control [101]. In this U.S. trial, participants in the automated condition on average logged in on 7.7 days and completed 5.6 out of 18 lessons (31.1% of the program). The authors reported a significant relationship between outcomes and adherence to the intervention, but baseline depression scores did not affect adherence. Significantly, the percentage of participants prescribed anti-depressants was relatively low in this study, being approximately one-third in both iCBT and TAU conditions. Also of note, in this trial participants were recruited through fliers in examination rooms, referral by physicians, and through letters sent to clinic patients. This recruitment method required potential participants to engage more pro-actively to enrol in the trial compared to the other two primary care trials where consent to participate was provided at the point of visiting the clinic. Accordingly, the study population in the U.S. trial may have been more motivated and willing to engage with an unguided technology intervention. This trial also screened for eligibility using a human-administered structured telephone interview (the M.I.N.I.) to assess depression status.

3.4.4 The use of iCBT in population with unmet need

Several studies in the review pointed to the successful implementation of automated iCBT for individuals whom it is likely had an unmet need for depression treatment [102, 105].

The trial by Lintvedt and colleagues demonstrated that automated iCBT can be effective for university students experiencing psychological distress [102]. This is of particular public health significance given the rate of depression in university students is higher than that of their age-matched counterparts in the general population [115] and that students report a range of barriers to accessing treatment including fear of stigma, fear of documentation of depression on their academic record, and cost [116]. Automated iCBT interventions can be accessed privately and at no cost to students. They can also be advertised widely on university campuses and implemented on a large scale at low cost, and so could be well suited for this setting. Further, such interventions may contribute to improved academic outcomes and associated career prospects and avert future episodes of depression. Priority should be given to further investigating the effectiveness of automated iCBT interventions in the university setting.

The results of the study by Farrer et al. [105] suggest that the provision of automated iCBT to telephone crisis helpline callers may be effective. The effect sizes for this trial were the largest of the studies included in the current review (*g* = 0.76 at post-test, *g* = 1.19 at 12-month follow-up) [105]. This could be because participants were highly depressed at baseline (CES-D: *M* = 36.4, *SD* = 10.17) and so had more opportunity to improve. However, it may also be that helpline callers are more receptive to distal treatment options. Individuals may be more motivated to engage with an intervention at a time of personal crisis. It is also possible that the callers preferred self-directed resources. Nevertheless, adherence to the intervention was relatively low.

By contrast with the trial involving university students, evidence from the sole school trial in the review suggests that automated iCBT may not be an effective intervention for universal prevention in a secondary school setting [100]. The fact that the vast majority of students in this trial did not once log into the program points to the need for some kind of teacher, counsellor or parental guidance in this population. Indeed, other studies with adolescents were screened out due to the presence of human supervision. Significantly, a previous universal prevention study in which the above iCBT program was delivered under the supervision of a teacher demonstrated its effectiveness in preventing depression among boys [117].

3.4.5 Adherence to automated iCBT interventions

As noted above, it is difficult to compare participant adherence to the automated iCBT interventions across studies due to variation in the information reported. However, it appears there are large differences both within and between interventions. For example, whereas only 8.5% of participants in a secondary student trial accessed *MoodGYM* [100], 58% of students in the university trial accessed the program at least once [102]. Similarly, in one community-based *Deprexis* trial, 46.2% of participants in the iCBT condition logged in at least three times [106], whereas in another 71.8% of participants logged in to complete at least four sessions [98].

Overall it appeared that in most studies only a minority of participants completed the entire iCBT intervention; however, this did not necessarily affect outcomes. Many trials failed to establish a dose-effect response, and it may be that completion of the entire intervention is not required in order to achieve significant reductions in depression. This is consistent with the results of a study that compared different combinations of *MoodGYM* modules and found that the first two modules of the program may be most important for improved depression outcomes in self-referred community users [118]. Further research is required to better understand the relative importance of different components of automated iCBT programs for improving depression symptoms.

It is important to note that many participants did not access the allocated iCBT intervention at all, suggesting that iCBT may not be a desirable intervention option for some people. There are a range of potential reasons that some individuals failed to access the iCBT program: for example, it is possible that at the time of enlisting in the trial they had a preference for face-to-face therapy or for an alternative treatment; there may have been technology barriers preventing them from accessing the program; or they may have lacked the motivation to undertake self-directed learning. Likewise, a large proportion of participants did not access the intervention more than once. Qualitative research has indicated that barriers to continued use of unguided iCBT can include: a desire for human support to maintain the user's motivation to continue using the program or to elaborate where needed on the concepts in the program; a lack of identification with the program content (not being able to apply the material to the user's personal situation); and technology access issues [119].

Although automated iCBT may not suit all individuals, either because they do not want to access it at all, or because they discover it does not suit them and they discontinue its use, the available evidence does support the potential impact of iCBT as a public health intervention. Many individuals do access the programs and experience positive results, and overall, as shown in the results of ITT analyses, in community settings there are consistent significant between-group differences in depressive symptoms for participants who are allocated to receive automated iCBT compared to those in control conditions.

3.4.6 Limitations

There are a number of limitations associated with the current review. First, the studies in the review employed a range of intervention programs of differing length and content. It is likely that the degree to which these interventions included CBT training or involved concepts other than CBT varied. However, there was more than one study for three interventions (*MoodGYM*, *Deprexis*, *Smiling is Fun*), allowing for a more direct comparison of outcomes and adherence rates across different settings. Moreover, the finding of positive results across different automated iCBT interventions suggests that the outcomes are robust and not dependent on a specific implementation of iCBT.

Second, most websites provide technical support if it is requested by the user, so it is possible that there was some minimal human contact in the included studies if users in the automated iCBT conditions requested technical assistance.

Third, there were limitations to the conclusions that could be drawn from comparing the magnitude of effect sizes across studies due to the difference in analyses used in the included studies. In particular, three studies did not analyse all randomised participants through ITT analysis (which

estimates the effect of the assigned treatment rather than the effect of the treatment that was received [120]). Instead, these studies reported the findings of completer analyses that only used data from participants who completed the relevant post-tests [95, 99, 100]. Another study used observed means to calculate the effect size (rather than means estimated from ITT analysis) [98]. In addition, several early studies imputed data within ITT analyses using the Last Observation Carried Forward (LOCF) method that assumes missing data is stable [104, 106, 107]. This method is no longer a preferred approach to dealing with missing data as it assumes that data is 'missing completely at random' (MCAR), with missing data not being related to participant or intervention characteristics in any way [121]. More recent studies employed linear mixed effect models for repeated measures containing both random and fixed effects, and which more appropriately assumes that data is 'missing at random' (MAR) [122].

There was also variation in the method used to calculate effect sizes. Where possible, effect sizes were extracted from the paper or computed by the current author, using the difference between the post-intervention time-point means for the control and the intervention conditions, divided by the pooled standard deviation at this later time-point. However, data was not always available to enable this calculation, and the original study authors were not contacted for this information. In addition, reported effect sizes were not always comparable – in two studies the reported effect sizes were based on the overall difference between the within-group pre-post effect sizes for the intervention and control conditions [107, 108], and in one study the effect size was calculated as the difference in mean change scores divided by the pooled standard deviation at baseline [94], which resulted in a larger effect size.

In addition, since the review was based on published trials only and trials with negative findings may be less likely to be published, it possible that the findings are subject to publication bias.

Finally, given that this review forms part of a post-graduate thesis, resource constraints precluded the use of multiple raters to evaluate studies against inclusion/exclusion criteria and to extract study data, and hence the assessments were undertaken by a single researcher (the candidate) only.

3.5 Conclusion

There is some evidence that fully automated (delivered with no human interaction) iCBT interventions can be effective in the reduction of depressive symptoms in users. In particular, the provision of automated iCBT to interested adults in the community appears to be effective for depression, regardless of whether individuals have a depression diagnosis. There is also preliminary

evidence of the effectiveness of iCBT for university students and crisis helpline callers. However, the usefulness of automated iCBT in primary care settings is less clear, and one study investigating unsupervised use with secondary students did not report positive results.

Further research using large community samples is required to investigate potential demographic predictors of positive depression outcomes for individuals choosing to access automated iCBT. More research is also required to understand why certain individuals respond to automated iCBT, including whether outcomes are related to a participant's motivation and treatment expectations at the outset. This has the potential to inform the decision-making of primary care physicians and other practitioners when determining to whom they should 'prescribe' automated self-help.

The cost of delivering automated interventions is relatively small compared to interventions that involve real-time human resources. Hence, given that there is evidence for effectiveness in improving depression symptoms in self-selected community users with varying baseline symptom levels, automated iCBT interventions could have a substantial public health impact if delivered worldwide to large numbers of community users at low or no cost to users [123]. As such, automated iCBT programs potentially provide an effective source of help for individuals who do not have access to other depression treatment modalities due to cost or availability, or who prefer to manage their mental health in a more private, self-directed way.

4. MoodGYM Spontaneous Community Users: User Characteristics

4.1 Introduction

As noted in previous chapters, the wide-spread, open-access delivery of automated iCBT programs for depression has the potential for a significant public health impact, given that: there is evidence for their effectiveness for reducing depression symptoms (particularly for users recruited from community settings), as demonstrated in Chapter 3; the cost of ongoing delivery of these interventions is relatively low compared to interventions which require real-time human input [124]; and the high prevalence and disability burden associated with depression is a global problem [1, 85, 86]. It is important to note that the findings from controlled efficacy trials may not translate into equivalent user outcomes in real-world users accessing interventions outside of research contexts [125]. The positive findings from a subset of trials of self-selected community participants accessing completely automated interventions provide credible evidence of the likely outcomes for real-world help-seeking community users [126]. Nonetheless, a detailed scientific study of the characteristics of and outcomes for real-world users who access self-help via automated iCBT is required to better understand the public health impact of such open access programs.

In this thesis, individuals who access publicly available interventions outside of research contexts are described as 'spontaneous community' users. This chapter describes the characteristics of the spontaneous community users of *MoodGYM*.

4.1.1 The literature on spontaneous iCBT users

Despite the large body of research on iCBT interventions, there is a lack of systematically documented information about the characteristics of or outcomes for users of globally accessible online programs for mental health.

In a recent systematic review, Fleming and colleagues examined observational or implementation studies which reported dissemination data from unguided digital interventions for depression or anxiety conducted outside of traditional research trial settings [127]. Eleven articles were identified, including six studies which involved iCBT interventions for depression[2]. Of these, five were studies of *MoodGYM* (Versions 1 and 2) [130, 131, 132, 133, 134], and one reported on a Spanish iCBT program

[2] An additional two studies were classified as treating both anxiety and depression using 'mixed' therapeutic approaches (CBT and positive psychology). In the current study these two interventions were not considered 'iCBT for depression', as one comprised apps (and was not Internet-based) [128] and the other ('Happify') included both a website and an app but focused on well-being as measured by the 'Happify Scale' rather than on depression [129].

'Help for Depression' (HDep) designed for female users [135]. The authors found no published RCTs or other effectiveness trials for the *HDep* intervention, and hence *MoodGYM* was the only depression iCBT program with effectiveness evidence for which published real-world observational studies were identified. However, it is informative to examine the characteristics of users of both programs since real-world users may be different from those included in research trials [127].

The *HDep* study considered 17,318 spontaneous community users who registered on the program and accessed the website at least twice over a 4-year period from March 2009 [135]. These registrants were predominately female (84.4%), and half (50.2%) were aged 18-30 years of age. Only 4.4% of these users were aged over 50 years. Most users were located in Mexico (95.0%) and were highly educated, with 64.8% having a university degree. The majority (63.5%) reported being single, and 60.9% were employed. Most registrants learnt about the program through Internet searches (83.2%). Depressive symptoms were assessed using the CES-D, and 97.1% of users who completed the scale (*n* = 16,564) had a score of 16 or more (clinically depressed cut-off score). In addition, users were asked about other mental health characteristics. More than half (54.9%) self-reported disability due to "feeling lonely, sad or listless", 37.8% had taken medication for "feeling nervous, anxious or excessively energetic" and approximately one third (30%) had previously received treatment for depression.

Given that *MoodGYM* is the most studied iCBT depression intervention in the 'real-world', and the subject of the current thesis, the published findings from the five studies investigating the characteristics of spontaneous *MoodGYM* community users are described in detail in the next section.

4.1.2 The extant literature on the characteristics of spontaneous MoodGYM community users

The *MoodGYM* program has been delivered worldwide since its launch in 2001 and was available free of charge to all individuals until July 2017[3]. Spontaneous visitors to the program are required to register a user account before they can access the program, and their registration data includes self-reported demographic information.

[3] From July 2017 individual users were required to pay an annual subscription fee to support ongoing delivery of the program. The program remained free for users from Australia, where delivery was funded by the Commonwealth Department of Health.

The user characteristics reported in studies of spontaneous *MoodGYM* community users are summarised in Table 4.1 and described in more detail below.

Table 4.1. User characteristics reported in studies of spontaneous MoodGYM community users

	Registration period	***N***	**Female (%)**	**Age**	**Goldberg Depression *M***	**Goldberg Anxiety *M***
Christensen H, et al., 2002 [130]	April 2001 – September 2001	2,909	60%	*M*=35.5 (13.0)	Male: 4.83 (95% CI: 4.57-5.09) Female: 5.33 (95% CI: 5.03-5.41)	Male: 5.27 (95% CI: 4.96-5.57) Female: 5.78 (95% CI: 5.54-6.02)
Christensen H, et al., 2004 [131]	April 2001 – September 2003	19,607	69.1% (n=9,873: completed 1 module)	NR	5.40 (*n*=9,873, *SD*=2.42)	5.60 (*n*=8,876, *SD*=2.51)
Christensen H, et al., 2006 [132]	April 2001 – October 2004	58,398 V1: *n*=19,607 V2: *n*=38,791	V1: 66% (*n*=19,607) V2: 60.8% (*n*=38,791)	NR	V1: 5.40 (*n*=9,873, *SD*=2.42) V2: 5.56 (*n*=26,822) *SD* NR	NR
Batterham PJ, et al., 2008 [133]	January 2006 – April 2007	82,159	66%	54% ≤ 35 years	NR	NR
Neil AL, et al., 2009 [134]	January 2006 – November 2007 (≤ 19 years)	7,207	72%	Included users ≤ 19 years)	5.46 (*SD*=2.42) *n* NR	5.50 (*SD*=2.59) *n* NR

NR= Not reported; V1= Version 1; V2= Version 2.

The first observational study of spontaneous *MoodGYM* users, published in 2002, reported on 2,909 community users who registered with the program over the first six months of its public delivery from April 2001 [130]. Most of these users of this Australian program were from the United States (34.9%) or Australia (33.2%). Over half (60%) of the registrants were female and the mean age of users was 35.5 years (*SD* = 13.0). Depressive and anxiety symptoms were measured with the nine-item Goldberg Depression and Anxiety scales [136], and the mean initial depression and anxiety scores were 4.83 (*n* = 406, 95% CI: 4.57, 5.09) and 5.27 (*n* = 280, 95% CI: 4.96, 5.57) respectively for

male users, and 5.22 (*n* = 690, 95% CI: 5.03, 5.41) and 5.78 (*n* = 406, 95% CI: 5.54, 6.02) respectively for female users. These scores were significantly higher than those in a random population sample of adults aged 20 to 24 years in Canberra, Australia (*N* = 2,385), suggesting that the level of depressive and anxiety symptoms among *MoodGYM* users was higher than that in the general population.

This data was also included in a subsequent study of 19,607 spontaneous users who registered on the *MoodGYM* site between April 2001 and September 2003 [131]. The paper reported only limited information about the characteristics of these users, with the study focusing on changes in user depression and anxiety scores. The mean initial Goldberg Depression and Anxiety scores were 5.40 (*n* = 9,873, *SD* = 2.42) and 5.60 (*n* = 8,876, *SD* = 2.51) respectively.

The third study of spontaneous users of *MoodGYM* included both those who registered on Version 1 of the program and an additional 38,791 community users who registered on Version 2 between September 2003 and October 2004 [132]. Version 2 of the program incorporated improvements to the research capability of the website, including the requirement for users to complete core assessments before progressing through the program. The study focused on predictors of final depression and anxiety scores, and comparisons of outcome for users of the different versions of the program. However, the research paper did note that the proportion of female registrants on Version 2 of *MoodGYM* was 60.8% (significantly lower than the 66% observed in the Version 1 dataset, $p < .001$) and that the mean initial depression score for Version 2 users was 5.56 (significantly higher than the mean initial score of 5.40 observed in Version 1, $p < .001$).

Two later studies of spontaneous users of *MoodGYM* focused on user adherence to the program but also included information about user characteristics [133, 134]. The first reported data from 82,159 community users of *MoodGYM* (Version 2) who registered during 2006 [133]. Of these users, 66% were female, 54% were less than 35 years of age, 22% were from a rural or remote area, and 18% reported being referred to the program by a mental health professional. Users had a high level of education, with 81% having completed secondary education and 40% having an undergraduate degree or higher degree. Most users were from Australia (41%) and the United Kingdom (31%). The second study focused on 7,207 younger public users who registered in 2006 or 2007 and self-reported to be aged 19 years or younger [134]. The majority of these users (72%) were female, 66% reported a history of depression and 19% lived in a rural or remote area. At baseline, the mean Goldberg Depression and Anxiety scores for these users were 5.46 (*SD* = 2.42) and 5.50 (*SD* = 2.59) respectively. The mean initial 'Warpy Thoughts Scale' score was also reported (*M* = 3.16, *SD* = 0.71). The 42-item Warpy Thoughts Scale measures levels of dysfunctional thinking [137] and is administered to *MoodGYM* users after baseline depression and anxiety surveys. Total scores were

calculated by taking the mean of the 5-point responses and range from 1 (lowest level) to 5 (highest level of dysfunctional thinking).

Limitations of existing studies

The existing observational studies of *MoodGYM* users were conducted early in the evolution of iCBT. Indeed the two earliest studies describe data collected in 2001-2003, which was also in the infancy of worldwide access and use of the Internet. Since these studies were published there has been an expansion of research trials in the domain (as noted in Chapter 2), increased access worldwide to technology and the Internet, the implementation of Australia's 'Telephone Counselling, Self Help and Web-Based Support Programmes' (Teleweb) government funding measure, and a likely increased awareness in the community of iCBT programs.

In addition, although community users accessed the program from a range of countries, the studies published to date did not undertake cross-country comparisons of user characteristics. This is significant because different countries are characterised by differing health service contexts and varied approaches to utilising the potential of e-mental health interventions. For example, in the United Kingdom computerised CBT (CCBT) has been included as part of the National Institute for Clinical Excellence (NICE) guidelines for the treatment of sub-threshold and mild to moderate depression in adults since 2002 [138]. Although not one of the recommended interventions in the guidelines, as an evidence-based CCBT program that could be accessed free to the end user, *MoodGYM* was widely promoted by National Health Service (NHS) trusts and is linked to by the U.K. NHS Choices health website. In Australia, the first e-mental health policy discussion paper was published in 2002 [139], and the public delivery of *MoodGYM* was supported with funding from the Australian Government from January 2007, and as part of the broader government Teleweb funding measure from July 2007. By contrast, to our knowledge, at this time e-mental health was not a focus of health policy, clinical practice guidelines, or government funding in the United States.

Another, although potentially related aspect not examined in any detail by the extant literature, is the referral of users to *MoodGYM* by health professionals. This is a key area of interest given that referral may reflect, at least in part, the level of integration and acceptance of e-mental health within different health systems. The referral of clients to an open-access website suggests a high level of knowledge and/or confidence in e-mental health resources by the referring health practitioner, which in turn may depend on e-mental related policies of and training opportunities available in different health systems. Hence an analysis of factors associated with health professional referrals may broaden our understanding of the implementation and uptake of iCBT as a public health intervention.

The studies to date also do not consider the motivation level of the user to engage with the program, nor the user's understanding of, or previous use of CBT. Additionally, there is limited consideration of the relationships between baseline depression and anxiety symptom levels and user characteristics.

4.1.3 Study aims

In view of the above limitations, the current study extends previous studies by considering a much larger period of public registrations in more detail. It has four aims:

Aim 1: *To examine the characteristics of spontaneous community users who registered on Version 3 of the MoodGYM program[4], over a period of approximately 5 years from March 2009*.

Aim 2: *To compare the demographic characteristics of spontaneous community users, and user pathways to the program, based on country of registration.*

Aim 3: *To investigate (i) differences in demographic characteristics of spontaneous community users who were and were not referred by a health professional, and (ii) whether referral could be predicted by these characteristics.*

Aim 4: *To investigate (i) the initial depression and anxiety symptom assessment scores of spontaneous community users, and (ii) the relationships between user characteristics and baseline symptom levels.*

4.2 Method

The study received approval from the Australian National University Human Research Ethics Committee (Protocol Number 2013/592).

4.2.1 Participants

Data were extracted from the *MoodGYM* database for users who registered with the Version 3 English-language release of the *MoodGYM* program between 6 March 2009 (version release date) and 6 January 2014.

[4] Version 3 of the *MoodGYM* program contained the same content as previous versions but was implemented on a new software platform to enable larger numbers of concurrent users and to enhance capabilities for research and translation.

Inclusion criteria

Data were included for all spontaneous users who created a user account in the English version of *MoodGYM* via submission of the registration web-form.

Exclusion criteria

Although Version 3 of the *MoodGYM* program was available to spontaneous users in several other languages (Norwegian, Chinese, Dutch and Finnish), the analyses in the current study were restricted to the English-language version.

Users were also excluded from the analysis if they accessed the *MoodGYM* program directly (without registration) as part of a specific research trial using a preloaded username and password.

Further, only users who registered for their own personal use were included in the analysis. The registration web-form incorporated the question:

> *"Which of the following reasons for visiting MoodGYM best applies to you?"* [I'm looking for help for myself/ I'm looking for help for someone else/ I'm a health professional who treats people with depression or anxiety/ I'm a researcher reviewing depression or anxiety sites/ I'm studying anxiety and depression as part of a college or university course].

Registrants could endorse more than one option. Only personal users – those who responded that they were seeking help for themselves – were incorporated in the analysis. Users who endorsed one or more of the remaining options only or who provided no response to the question were excluded from the study.

Finally, users identified as test user accounts created by staff at the Australian National University were also excluded.

4.2.2 Measures

Self-report data from the registration form were collected for each user together with a record of the date and time of registration. The registration form included questions about the user's demographic characteristics, the pathway through which they found the program, their depression and treatment history, and the degree to which they were motivated to complete *MoodGYM*.

Registration data

Gender and age: Users were asked to indicate their gender [male / female] and age category [15 years or under / 16-19 years / 20-24 years / 25-29 years / 30-34 years / 35-39 years / 40-44 years / 45-49 years / 50-54 years / 55-59 years / 60-64 years / 65-69 years /70-74 years / 75 years or older]. These questions were compulsory.

For the purpose of analysis, the 14 age categories were recoded into seven broader categories [15 years or under / 16-19 years / 20-29 years / 30-39 years / 40-49 years / 50-59 years / 60 years or older].

Country of registration: Users were asked the optional question "*Which country are you logging in from?"* and provided with a drop-down list of 239 countries and subdivisions generated from the ISO 3166 International Standard for country codes. This was used to categorise the user's country of residence (rather than, for example, the IP address from which they are accessing the program).

Location: Users were required to answer the mandatory question, "*Are you in a city area or in a rural/remote region?*" [Capital city / Other city / Rural or remote region].

Education level: User education level was assessed at registration through two compulsory questions:

1) *"What is the highest level of schooling that you have completed?"* [Some primary / All of primary / Some secondary / Four years of secondary / All of secondary]; and

2) *"What is the highest level of post-secondary/tertiary education you that you have completed?"* [None / Trade or apprenticeship / Other certificate / Current undergraduate / Diploma / Bachelor's degree / Higher degree / Other].

The education data were subsequently recoded into the following seven categories:

- Some or no secondary schooling (users who did not check 'All of secondary' at question 1, and who checked 'none' at question 2).
- Secondary schooling (users who checked 'All of secondary' at question 1, and who checked 'none' at question 2).
- Trade, apprenticeship or certificate (users who checked 'Trade or apprenticeship' or 'Other certificate' or 'Other' at question 2).
- Current undergraduate student (users who checked 'Current undergraduate' at question 2).

- Diploma (users who checked 'Diploma' at question 2).
- Bachelor's Degree (users who checked 'Bachelor's Degree' at question 2).
- Higher degree (users who checked 'Higher degree' at question 2).

Pathway to *MoodGYM*: The optional question "*How did you find out about the MoodGYM site?"* was included along with the following response options:

- I was referred by my GP/doctor
- I was referred by my counsellor or psychologist or psychiatrist
- I was referred by a friend
- I was referred by a family member
- I was referred by a work colleague or supervisor
- I was referred by my teacher or lecturer
- I searched for online CBT using a search engine
- I searched for information about depression using a search engine
- I clicked on a link from another website
- I read about the program in the media
- Other

For the purpose of analysis the categories of 'I searched for online CBT using a search engine' and 'I searched for information about depression using a search engine' were combined into one 'Internet search' category, and the responses 'I was referred by a friend' and 'I was referred by a family member' were combined into one 'family/friend' category.

Previous depression and treatment history: Self-reported previous history of depression was assessed using the following compulsory question "*Have you ever in your life been markedly depressed? (That is for several weeks or more, you felt sad, lost interest in things and felt lacking in energy)"* [Yes / No].

Past help-seeking and treatment were assessed using the item: "*Have you used any of the following to help yourself cope with depression? (You may choose more than one)"* with the following response options comprising evidence-based treatments for depression:

- GP
- Psychiatrist
- Psychologist
- Other mental health professional
- St John's Wort
- Anti-depressants
- Exercise
- Light therapy
- Self-help books
- Electroconvulsive therapy (ECT)
- Interpersonal therapy
- None of the above

Registrants were also asked about their knowledge and previous use of cognitive behaviour therapy: *"Have you ever used Cognitive Behaviour Therapy (CBT) to help yourself cope with depression?"* [Yes, I found it useful / Yes, but I didn't find it useful / No, but I want to try it / No, and it's not something that I want to try / I'm not sure what CBT is].

An evidence-based treatment score was calculated based on the total number of evidence-based treatments reported to have been used by each user. The latter included the number of treatments endorsed in the multiple option item plus any affirmative response to the CBT usage item.

Motivation level: Motivation to complete *MoodGYM* was assessed with the following single item: *"Which of the following statements best applies to you?"* There were three response options:

- I am committed to working through the entire *MoodGYM* program
- I'm not sure what MoodGYM involves, but I'll try it out
- I don't think I'll work through the entire program, but I'm interested in having a look

Severity of baseline depression and anxiety: The user's baseline anxiety and depressive symptoms were assessed following registration and prior to the commencement of the first *MoodGYM* module using the 9-point Goldberg Depression and Anxiety scales – self-report scales with evidence of adequate reliability and validity [136]. For each of the scales, items are scored 0 (no) or 1 (yes) and

summed. The maximum score of 9 represents very high depression [anxiety] symptoms, while a score of 0 represents no depression [anxiety] symptoms.

4.2.3 Analyses

All statistical analyses were conducted using SPSS Version 25.0 [140].

Aim 1: User characteristics

The frequency of MoodGYM registrants (personal users) was computed for each month between April 2009 and December 2013. In order to investigate the trend of registration numbers over time, taking into account monthly seasonal variations, a centred 12-month moving mean was computed and a linear regression model fitted to this data. The individual monthly seasonal effects were calculated as the difference between the centred moving mean and the raw number of registrations for each month, and an average seasonal effect was calculated for each month.

Descriptive statistics were calculated for each of the demographic characteristics (age, gender, location, education, country of registration), for user motivation level, and for the pathway through which users found *MoodGYM* over time. The age distributions of male and female users were compared using Mann-Whitney tests.

Aim 2: Characteristics of users from different countries

Sub-analyses were undertaken to compare the demographic characteristics of registrants, number of 'motivated' users and pathways to accessing *MoodGYM* across of the four countries with the highest user registrations. Comparisons were undertaken using overall Chi-square tests incorporating all four country categories followed by pairwise post hoc tests using a Bonferroni correction to correct for the type 1 error associated with multiple comparisons. For the purposes of this analysis, the categories 'online search' and 'link from a website' in the 'pathway to MoodGYM' variable were merged into a single category, as were the categories 'colleague/supervisor' and 'teacher/lecturer'.

Aim 3: Referral by a health professional

In order to test whether referral to the program by a health professional could be predicted using demographic data, a logistic regression was conducted. The dependent variable was 'Referred by a health professional' (Yes, No), coded 'Yes' if the user endorsed either 'I was referred by my GP/doctor' or 'I was referred by my counsellor or psychologist or psychiatrist' when asked how they found the site.

The independent variables included in the model were: gender, age group, country, location, motivation response and education level. The youngest and oldest age group categories were further collapsed into five age groups for the analysis (19 years or younger; 20-29 years; 30-39 years; 40-49 years; 50 years or older). Location was coded into two categories (rural/remote region or not), based on whether the user responded 'Rural or remote region' to the question about their location. 'Motivated' (Yes, No) was coded 'Yes' if users responded, 'I am committed to working through the entire MoodGYM program' to the motivation question. All independent variables with more than two categories were dummy coded for ease of interpretation of odds ratios in the logistic regression model.

Independent variables were entered into the model in two blocks: gender, age group, education level; country, location, motivated. Hierarchical analysis was chosen so that the relative contribution of country, location and motivation could be examined, after controlling for other demographic characteristics.

Aim 4: Baseline depression and anxiety scores

The mean Goldberg depression and anxiety scores for participants who completed the initial symptom scales were examined, and exploratory analyses were undertaken to compare mean depression and anxiety symptom scores across participant sub-groups.

Sub-group comparisons: univariate analyses

The initial symptom levels for variables that involved two sub-groups were compared using independent *t*-tests. The dichotomous variables included: (female) gender, rural location, motivation status and health professional referral (all coded Yes, No, as per the previous analyses), as well as 'History of depression' (Yes, No) – coded 'Yes' if users reported that they had ever in their life been markedly depressed. In each case, symptom data distributions, boxplots and P-P plots were examined, and normality tests undertaken. Although the assumption of normality was violated for all variables (Kolmogorov-Smirnov tests significant to <.001), *t*-test comparisons were selected as appropriate given the large sample size [141].

Where a variable involved more than two categories (age group and education level), between-group differences in initial symptoms levels were examined using one-way between-subjects ANOVAs. For each analysis, a Levene's test was undertaken to test if the assumption of homogeneity of variance had been violated. As expected, given the large sample sizes involved, the assumption was violated in each case. Given this and the inequality of sub-group sizes, Welch's *F* were computed for each

analysis. Games-Howell post hoc comparison tests were also conducted to explore differences between groups.

In order to explore the interaction effect of gender by age group on symptom scores, a gender x age variable was coded, with ten categories for each age/gender combination. One-way ANOVAs were then performed using age x gender as the independent variable. Games-Howell post ho*c* comparison tests were also conducted, with comparisons between male and female users examined within each age group.

Mean depression and anxiety scores were also examined as a function of the number of treatments for depression previously tried by users (continuous variable), and a bivariate linear regression analysis was conducted to explore the relationship between the number of prior treatments and symptom scores.

Predicting baseline depression and anxiety symptom scores: multivariate analyses

Separate standard multiple regression analyses were conducted to test whether initial Goldberg depression and initial anxiety scores (dependent variables) could be predicted from users' registration data (independent variables).

The independent variables included in the models were:

- Gender: female [0, 1];
- Age group: Dummy coded with ≥ 50 years as the references group, ≤ 19 years [0, 1], 20-29 years [0, 1], 30-39 years [0, 1], 40-49 years [0, 1];
- Country: Dummy coded with Australia as the reference group, U.K. [0, 1], Canada [0, 1], U.S. [0, 1];
- Location: rural [0, 1];
- Education level: Dummy coded with Higher degree as the reference group, Some/no secondary [0, 1], Secondary only [0, 1], Trade/certificate [0, 1], Undergraduate [0, 1], Diploma [0, 1], Bachelor's degree [0, 1].
- Referral by a health professional: referred [0, 1];
- Previous history of depression: depression history [0, 1];
- Number of treatments previously tried: treatments [scale from 0 to 12];
- Motivated to complete the program: motivated [0, 1].

For both analyses assumptions of linearity, independence of errors, homoscedasticity and normality of residuals were checked. In addition, Variance Inflation Factors (VIF) were examined for the independent variables in each model to identify correlations and ensure that multicollinearity was within acceptable levels.

4.3 Results

4.3.1 Aim 1: User characteristics

During the period between 6 March 2009 and 6 January 2014, 417,354 eligible spontaneous users registered with the Version 3 English language release of *MoodGYM*. An additional 65,753 users who registered during this time period were excluded because they were not personal users. Rather, they fell into one or more of the following categories: a person seeking help for someone else (*n* = 38,460), a professional who treats people for depression or anxiety (*n* = 30,434), a researcher reviewing depression sites (*n* = 8,788), or a student studying anxiety and depression as part of a college or university course (*n* = 17,693). In addition, 53 test user accounts were removed from the dataset.

Of the eligible participants seeking personal help and thus included in the analysis, 7.1% (*n* = 29,432) endorsed an additional reason for visiting *MoodGYM*, the majority of whom were looking for help for someone else as well as themselves (see Table 4.2).

Table 4.2. Additional reason for visiting MoodGYM (eligible participants, N=417,354).

Reason	**n (%)**
I'm looking for help for someone else	19,622 (4.7%)
I'm a health professional who treats people with depression or anxiety	7,571 (1.8%)
I'm a researcher reviewing depression or anxiety sites	2,298 (0.6%)
I'm studying anxiety and depression as part of a college or university course	5,644 (1.4%)

Registration over time

Figure 4.1 displays the number of eligible personal user registrations on *MoodGYM*, for each month from April 2009 to December 2013 (*n* = 409,111). Users who registered during March 2009 (*n* = 6,920) and January 2014 (*n* = 1,323) are not included in Figure 4.1 since data were only collected

for 25 and 6 days of these months respectively. On average there were 7,177 new registrations each month.

A linear regression analysis of the 12-month centred moving means over time showed a significant effect of time on registration numbers adjusted for seasonal variation ($F(1,43) = 213.31$, $p < .001$), with an adjusted R^2 of 0.828. The frequency of new registrations increased by 26.4 registrations for each month, plus or minus the average seasonal adjustment for that month of the year (as shown in Table 4.3). The largest negative seasonal effect was found for December (minus 1,406 registrations) and March and May had the largest positive seasonal effects (plus 630 and 603 registrations respectively).

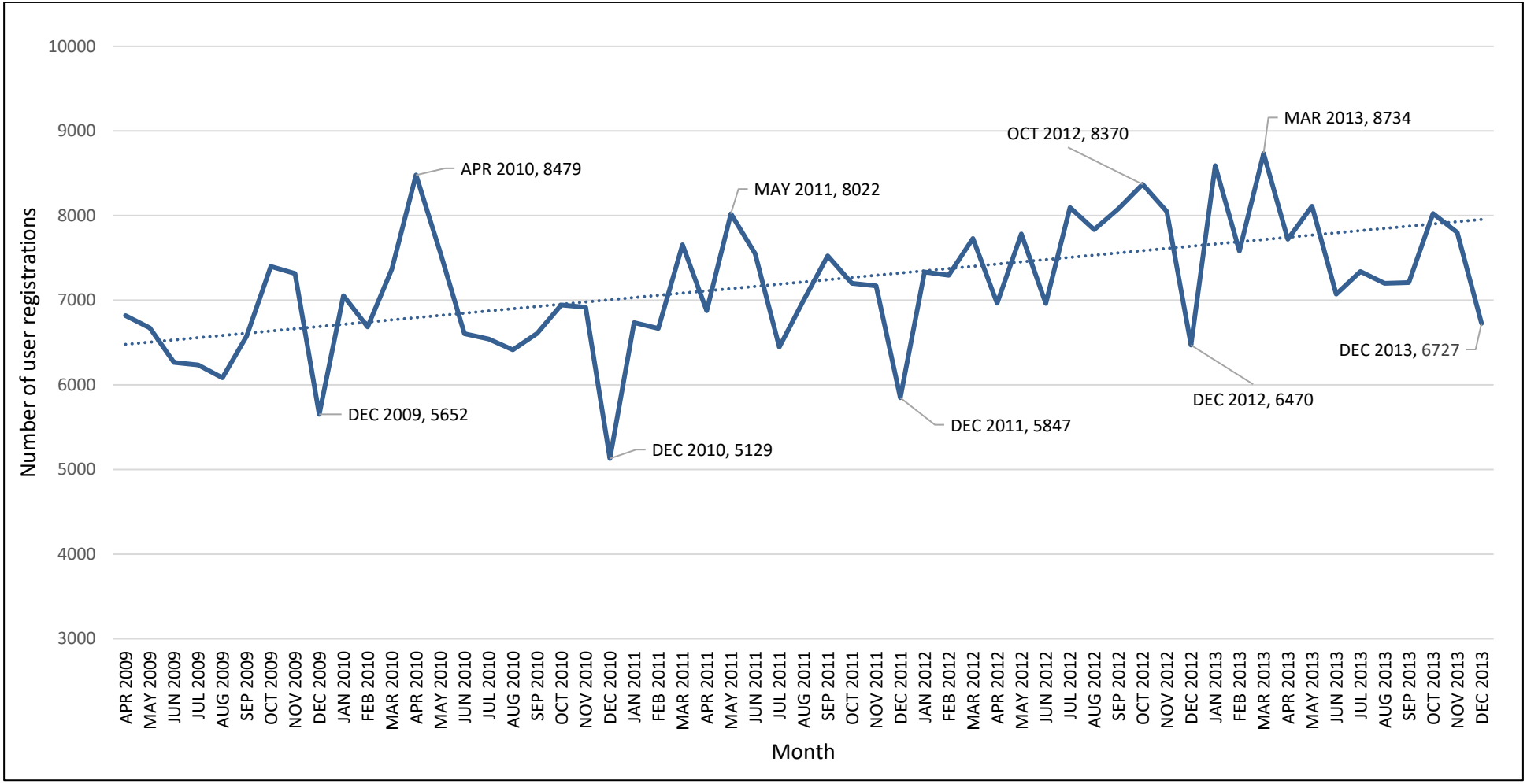

Figure 4.1. Number of new user registrations (eligible participants), by month (n=409,111).
Data collected during March 2009 and January 2014 not included (n=8,243).
Trendline represents linear regression for centred moving mean (y=26.364x +6451.7).

Table 4.3. Average seasonal effect on new registration numbers, for each month of the year.

Month of the year	Average seasonal effect
January	229
February	-166
March	630
April	255
May	605
June	-234
July	-158
August	-140
September	152
October	346
November	207
December	-1406

Age and gender

Of the 417,354 eligible participants, the majority were female (*n* = 284,148, 68.1%); the modal age was between 20-39 years (*n* = 130,503, 31.3%). The age group distribution of male and female users is displayed in Figure 4.2.

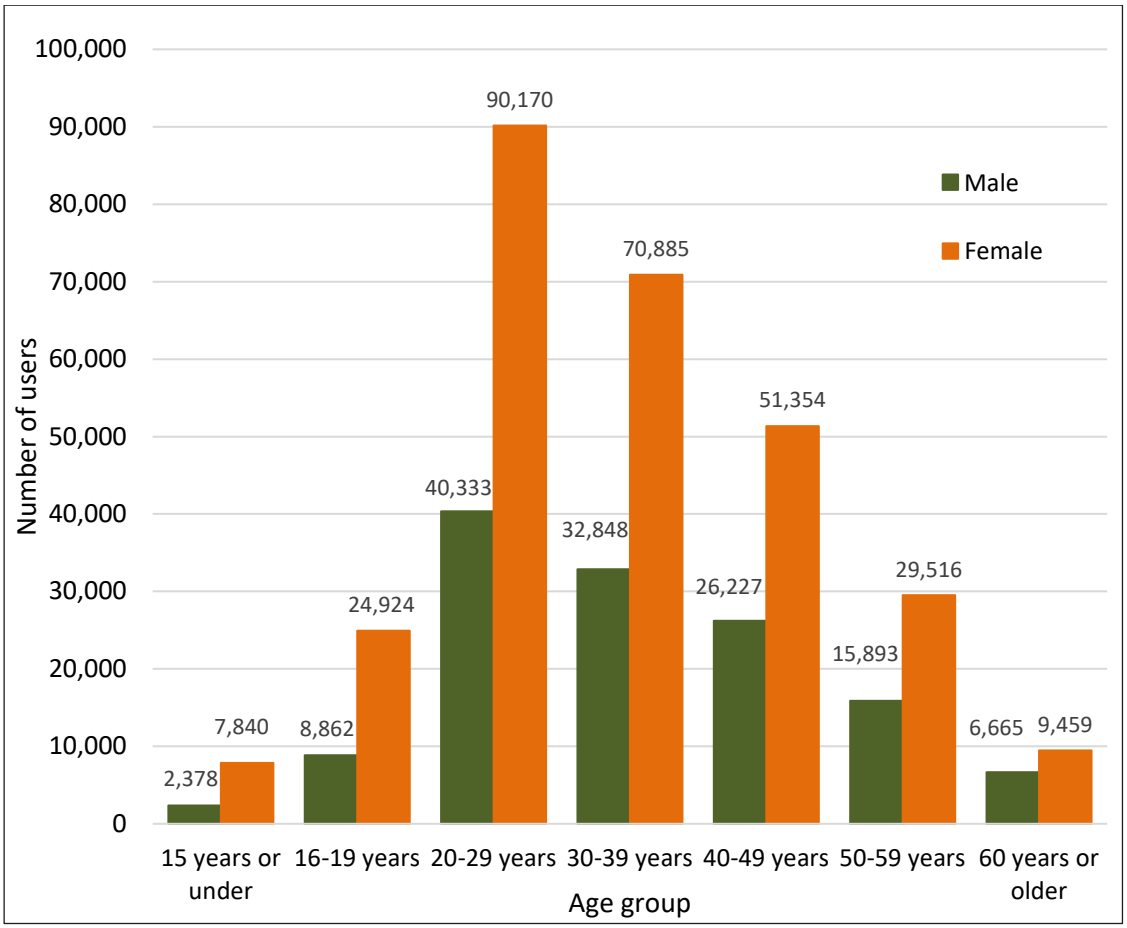

Figure 4.2. Age group and gender of MoodGYM users (N=417,354).

Although similar, the age distributions differed somewhat for male compared to female users. The modal frequency for male and female users was the same (20-29 years age group). However, a Mann-Whitney U test indicated that male users were older than female users ($U = 1.75 \times 10^{10}$, $z = -38.44$, $p < .001$) with visual inspection indicating that the relative number of female users decreased more sharply in the older age groups than for male users. This trend is clearly illustrated in Figure 4.3, which shows that the percentage of male users within each age group increased with age. Only 23.3% of users aged 15 years or under were male, compared to 30.9% of those aged 20-29 years, and 41.3% of users 60 years or older.

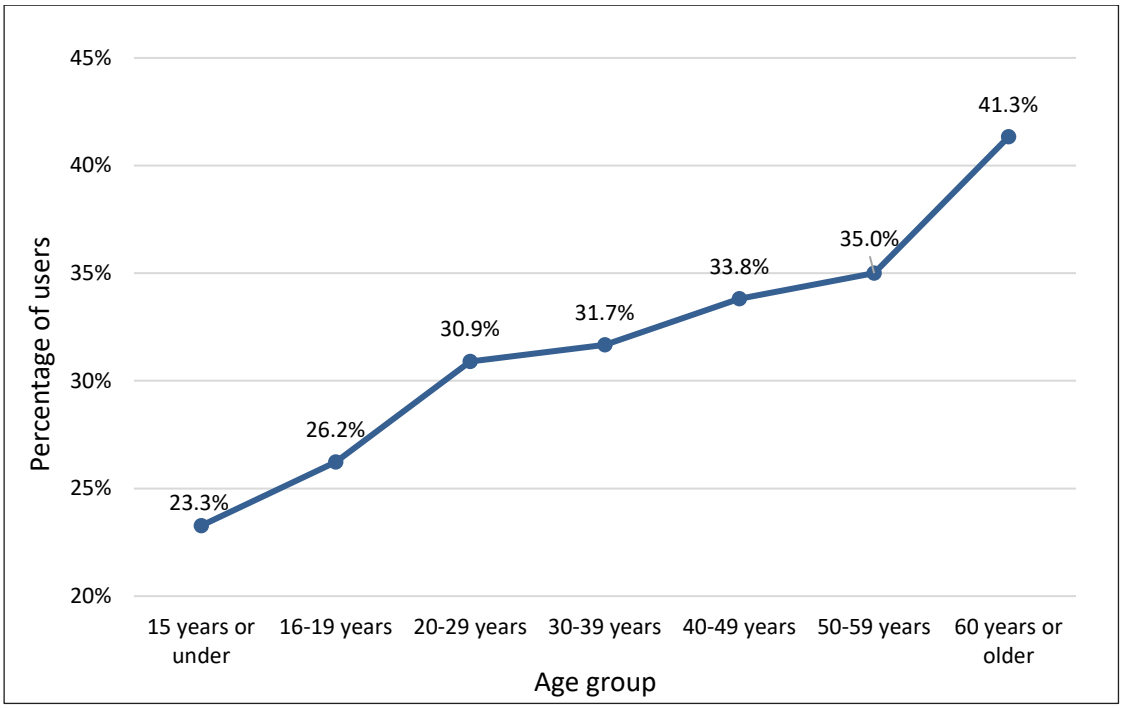

Figure 4.3. Percentage of MoodGYM users within each age group who were male (N=417,354).

Country of registration

Table 4.4 shows the frequencies of users registering from each country. The largest percentage of users reported logging in from the United Kingdom (49.5%), followed by Australia (20.8%) and Canada (6.5%). Of the eligible participants, 38,103 (9.1%) did not respond to this non-mandatory question.

Table 4.4. Country of registration ("Which country are you logging in from?", N=417,354).

Country	**N (% total)**
United Kingdom	206,698 (49.5%)
Australia	86,676 (20.8%)
Canada	27,109 (6.5%)
United States	19,830 (4.8%)
New Zealand	8,809 (2.1%)
Ireland	6,213 (1.5%)
No response	38,103 (9.1%)
Other country (combined)	23,916 (5.7%)
Total	**417,354 (100%)**

Location

Approximately one quarter (102,499, 24.6%) of users reported residing in a rural or remote region. 27.9% (116,631) resided in a capital city, and 47.5% (198,224) responded that they were located in a non-capital city.

Education level

The educational levels of personal users are shown in Figure 4.4. The majority (82.3%) of personal users indicated that they had undertaken post-secondary education, with 40.5% of users having completed a university Bachelor's or higher degree.

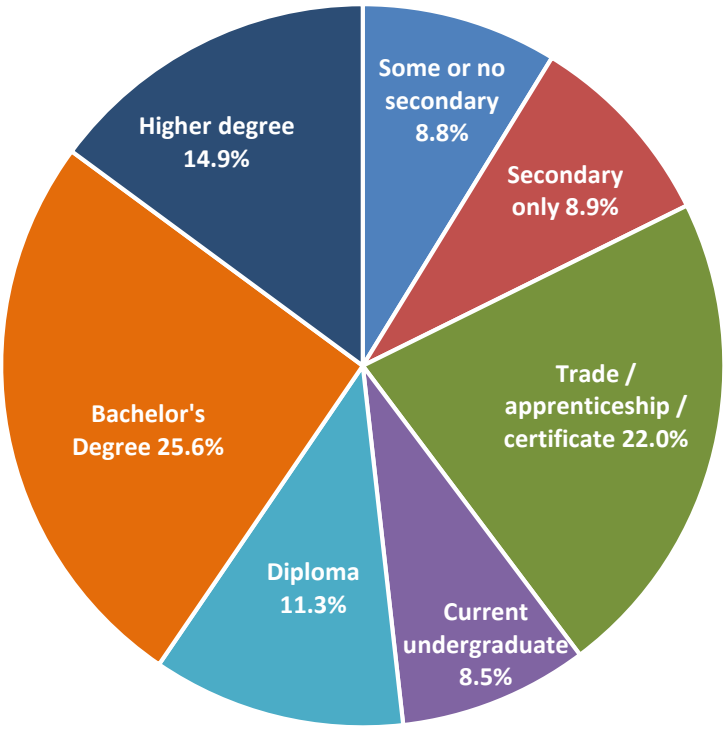

Figure 4.4. Education level of MoodGYM users (N=317,354).

Pathway to MoodGYM

Figure 4.5 depicts the methods that registrants employed to locate the program. A total of 42.7% of personal users indicated at registration that they were referred to the site by a health professional (GP, doctor, counsellor, psychologist or psychiatrist), 35.0% that they found the site via Internet search, website links or news stories, and 14.2% that they were referred by a friend or family member.

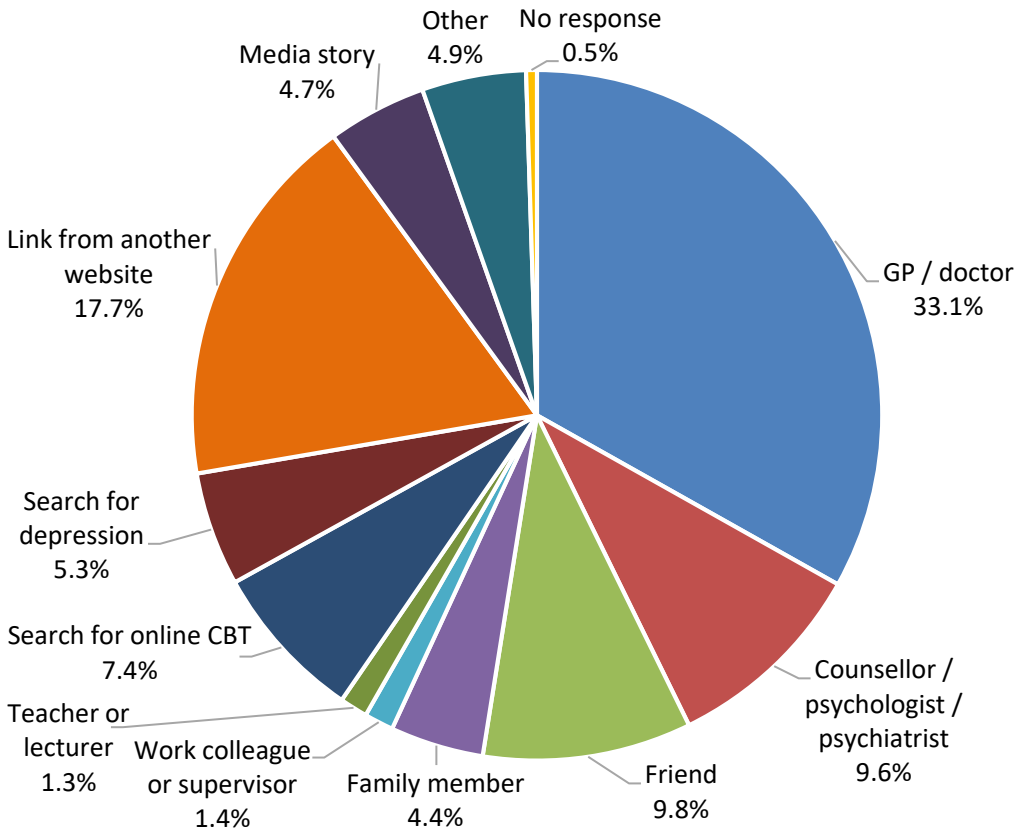

Figure 4.5. How personal users found the MoodGYM program (N=417,354).

The numbers of users finding *MoodGYM* through different routes were also examined by month of registration. As described above, user data from March 2009 (*n* = 6,920) and January 2014 (*n* = 1,323) were excluded since data was not collected for part of these months. Figure 4.6 shows the number of users who reported finding *MoodGYM* via each method, by month of registration (for each full month from April 2009 to December 2013, *n* = 407,030). Over 2,000 users who registered during this time period did not provide a response and were not included.

There are fluctuations in user numbers across all methods of locating *MoodGYM*. However, it is evident that prominent peaks in user registrations occurred for the media ("I read about the program in the media") at specific time-points. Similarly, there are pronounced peaks in the "I clicked on a link from another website" category at some time-points.

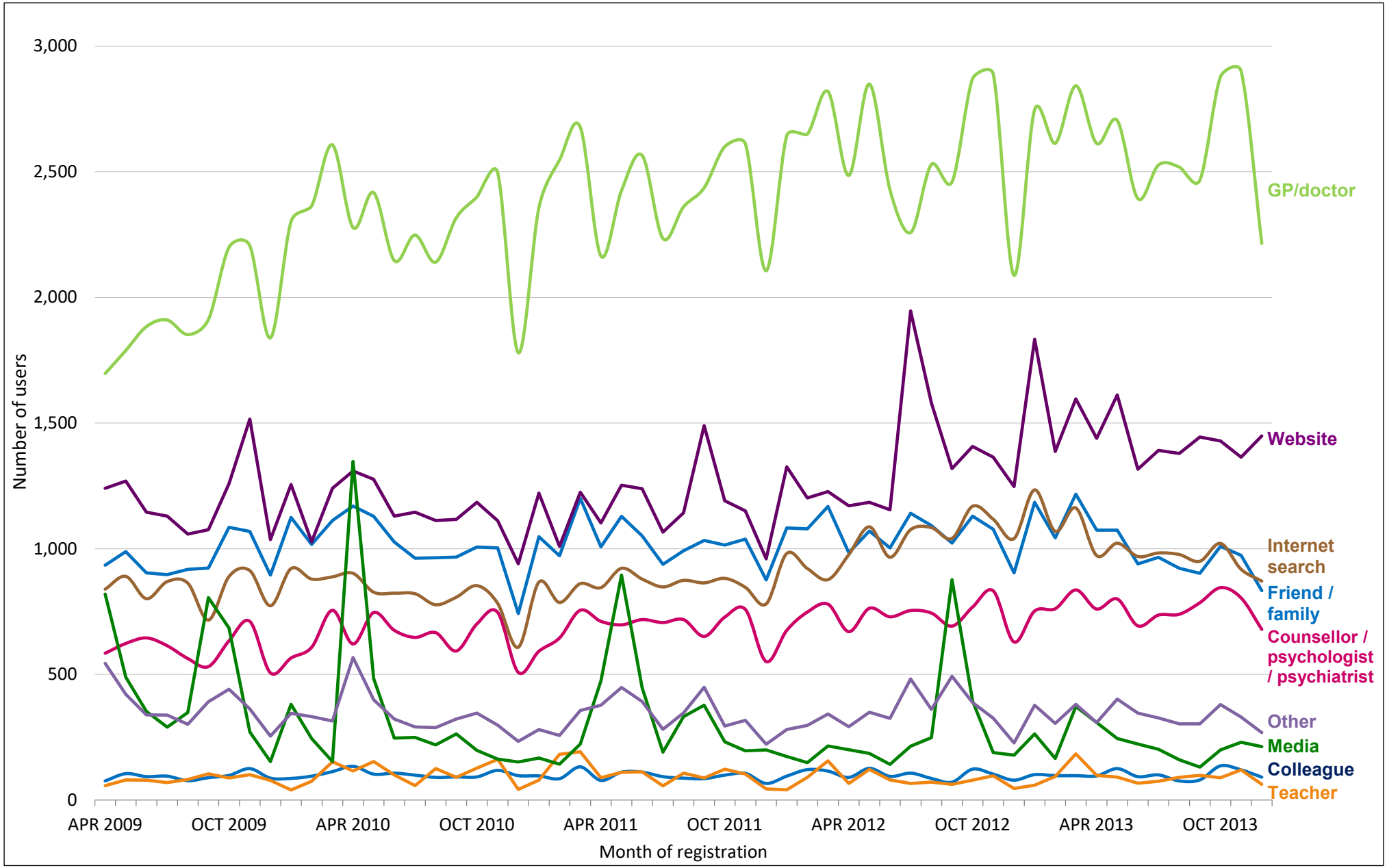

Figure 4.6. How users found MoodGYM, by month of registration n=407,030.
(No response - not included, n=2,081).

Previous depression and treatment history

Most users (352,781, 84.5%) reported that they have previously been depressed. Even more users (368,530, 88.3%) indicated having used at least one evidence-based treatment option for coping with depression.

Figure 4.7 shows the number of evidence-based treatments for depression endorsed by each user at registration. Approximately 88% had previously accessed at least one treatment with the majority of users (57.6%) indicating that they had used three or more of the options.

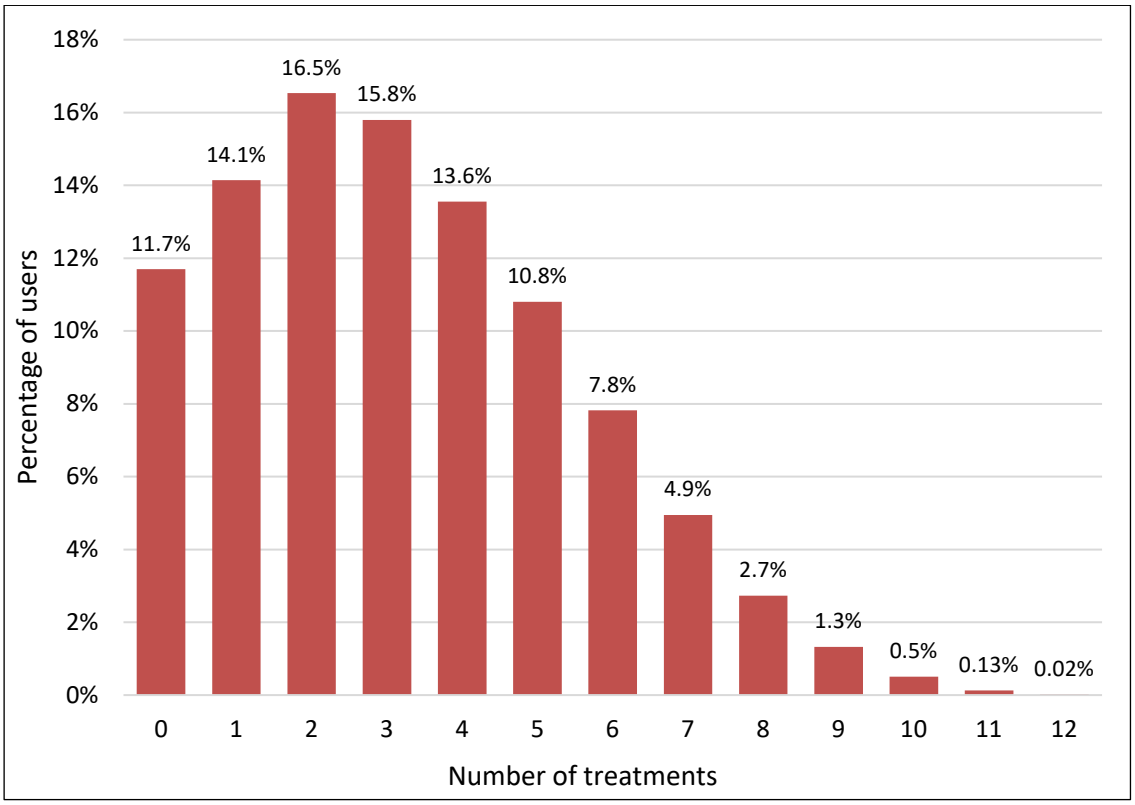

Figure 4.7. Self-reported number of evidence-based treatments for depression used by MoodGYM users (N=417,354).

Figure 4.8 shows the percentage of users who indicated that they had accessed each form of help. The most commonly used treatment sources were anti-depressant medication (50.6% of users), exercise (49.0% of users), and visiting the general practitioner (37.8%).

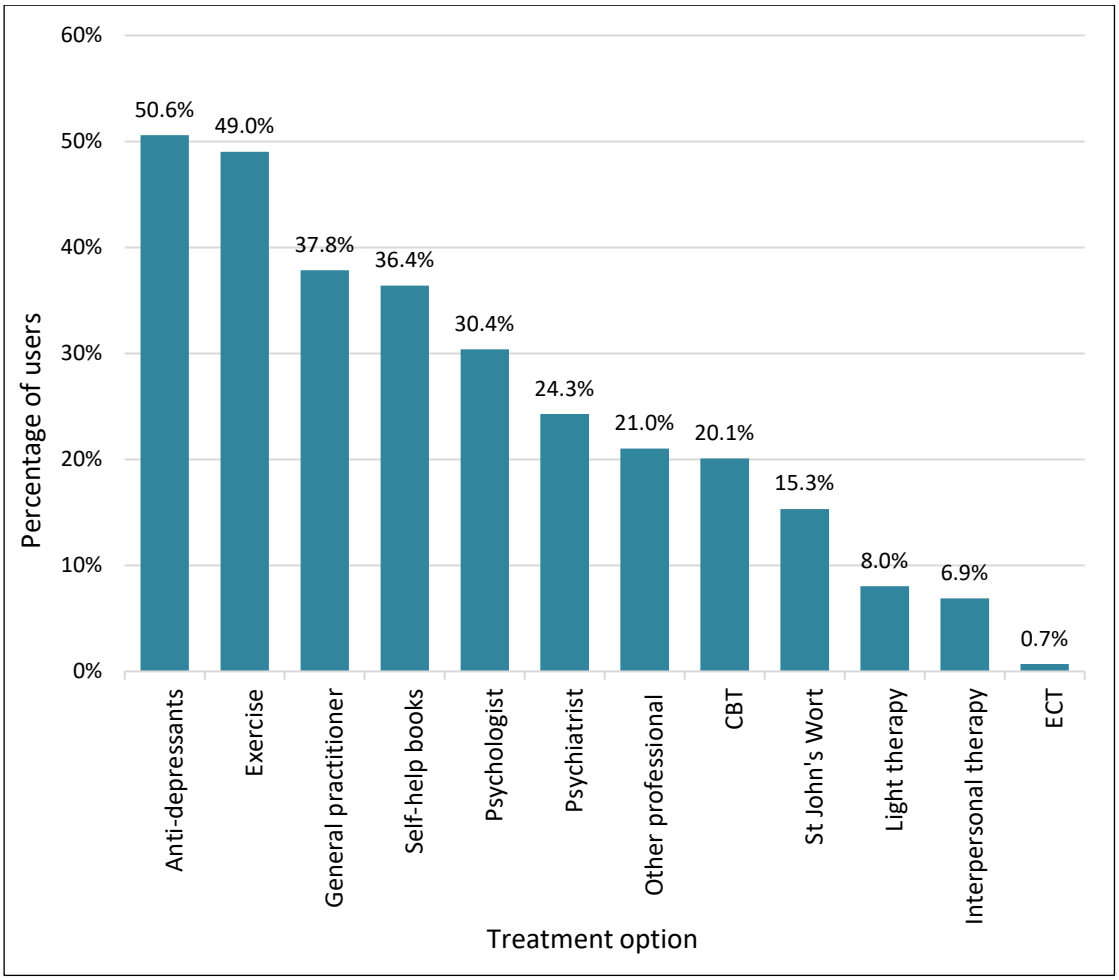

Figure 4.8. Percentage of MoodGYM users who endorsed using each treatment option to cope with depression (N=417,354).

Approximately three quarters (75.9%) of users reported that they had sought help from a health professional for depression (GP, psychologist, psychiatrist or other health professional). More than half (57.4%) indicated that they had previously tried a 'conventional' depression treatment (anti-depressants, CBT, Interpersonal therapy or ECT), and 63.8% had used an alternative or lifestyle treatment for depression (exercise, self-help books, St John's Wort or light therapy).

Frequencies of user responses to the question "*Have you ever used Cognitive Behaviour Therapy (CBT) to help yourself cope with depression?"* are shown in Figure 4.9. A total of 83,892 (20.1%) of the users indicated that they had previously used CBT to cope with depression. Of these 71.0% reported finding it useful. The majority of users (52.1%) indicated that they had not used CBT but wanted to try it.

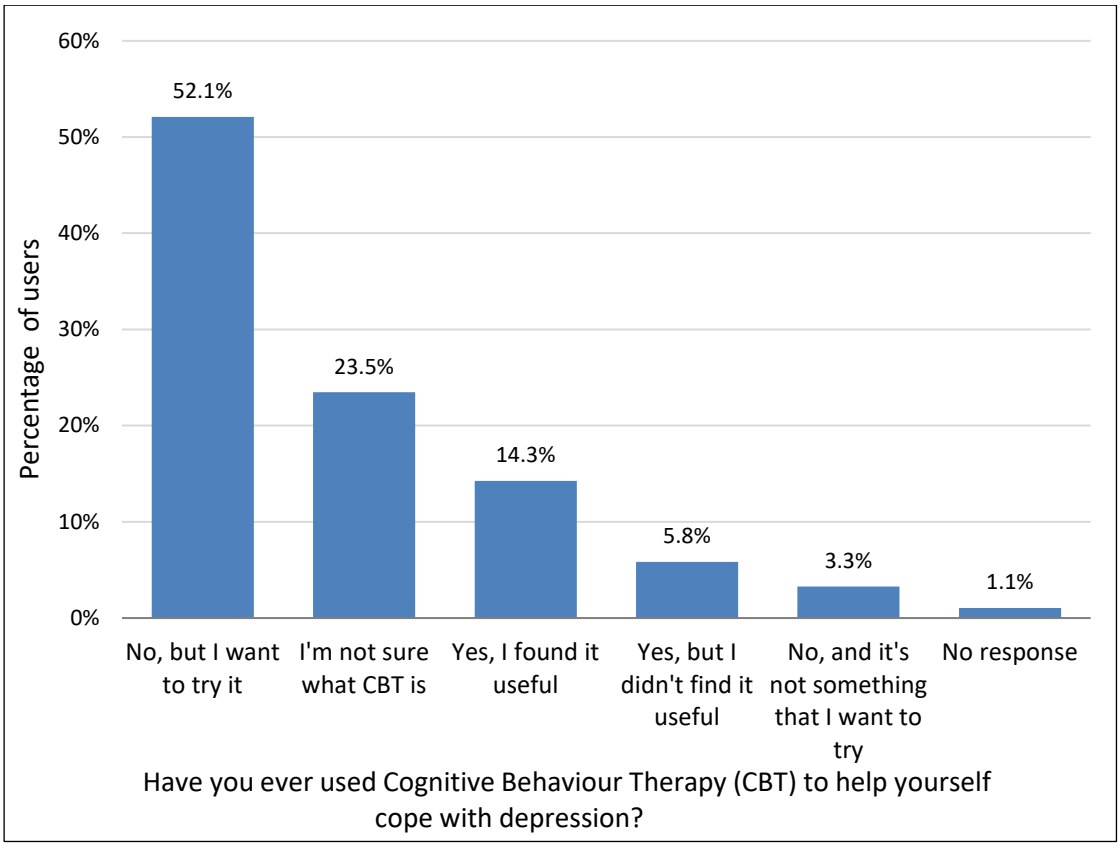

Figure 4.9. User responses to whether they have used CBT to help cope with depression (N=317,354).

Motivation level

When asked which of the statements in the Motivation item best applied to themselves, 71.9% of personal users selected: "*I'm not sure what MoodGYM involves, but I'll try it out*" (*N* = 417,354). User response frequencies are shown in Table 4.5.

Table 4.5. MoodGYM user responses to the Motivation item (N=417,354).

	N (% total)
I'm not sure what MoodGYM involves, but I'll try it out	300,099 (71.9%)
I am committed to working through the entire program	103,266 (24.7%)
I don't think I'll work through the entire program, but I'm interested in having a look	12,715 (3.0%)
No response	1,274 (0.3%)
Total	**417,354 (100%)**

4.3.2 *Aim 2: Comparisons by country*

Gender

Figure 4.10 displays the number of male users logging in from the four countries with the highest number of user registrations (n = 340,313). The relationship between the country and gender of users was statistically significant, χ^2 (3, n = 340,313) = 441.15 p < .001), but weak, Cramer's V = 0.036.

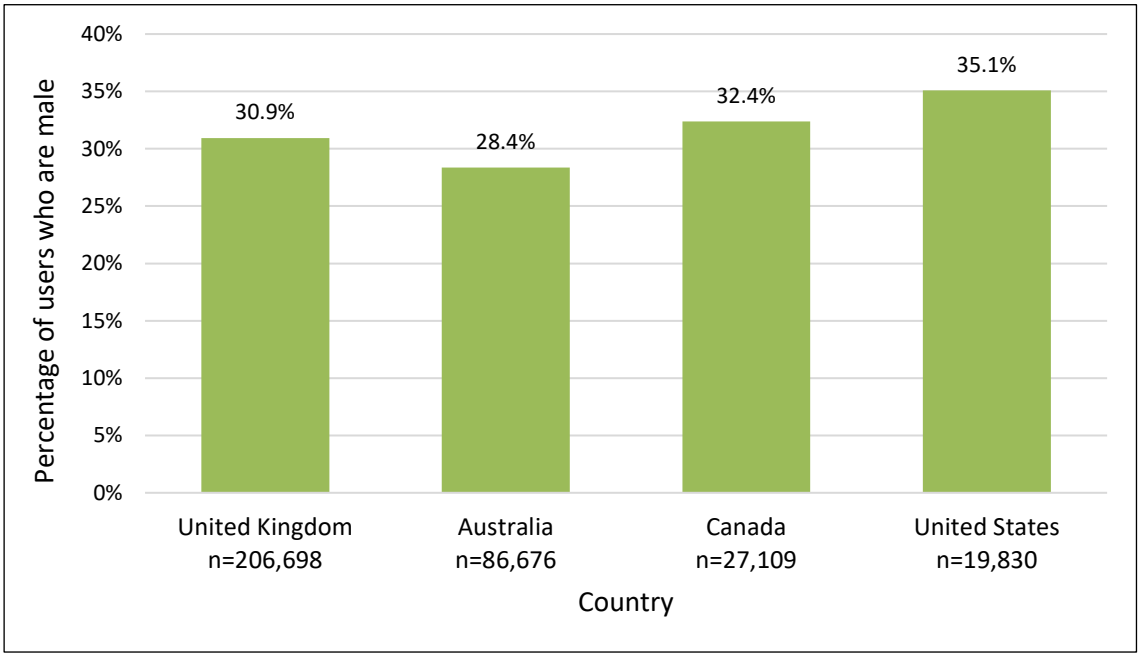

Figure 4.10. Percentage of users who are male, by country (n=340,313).

Post hoc analyses were also undertaken to investigate the differences in gender between each pair of countries. Chi-square tests, using Bonferroni corrections to account for the six comparisons, showed statistically significant relationships between gender and country for all pairs (see Table 4.6). In particular, the percentage of males among U.S. users was higher than in the other countries, there were a greater proportion of male users among the Canadian than the Australian and U.K. visitors, and the proportion of males was greater among Australian than the U.K. users.

Table 4.6. Chi-square test statistics: Gender by pairwise country comparisons

Comparison	**$\chi^2(1)$**	***n***	***p***	**Bonferroni adjusted *p***	**Cramer's *V***
U.K. vs Australia	191.07	293,374	<.001	<.001	0.026
U.K. vs Canada	23.00	233,807	<.001	<.001	0.010
U.K. vs U.S.	144.52	226,528	<.001	<.001	0.025
Australia vs Canada	159.37	113,785	<.001	<.001	0.037
Australia vs U.S.	348.75	106,506	<.001	<.001	0.057
Canada vs U.S.	37.75	46,939	<.001	<.001	0.028

Age group

The age group distributions for users logging in from these four countries were also compared (n = 340,313). Figure 4.11 shows the distributions for users from each country. A Chi-square test of independence showed a statistically significant relationship between country and age group, χ^2 (18, n = 340,313) = 7,287.70, p < .001 although this relationship was not strong, Cramer's V = .084.

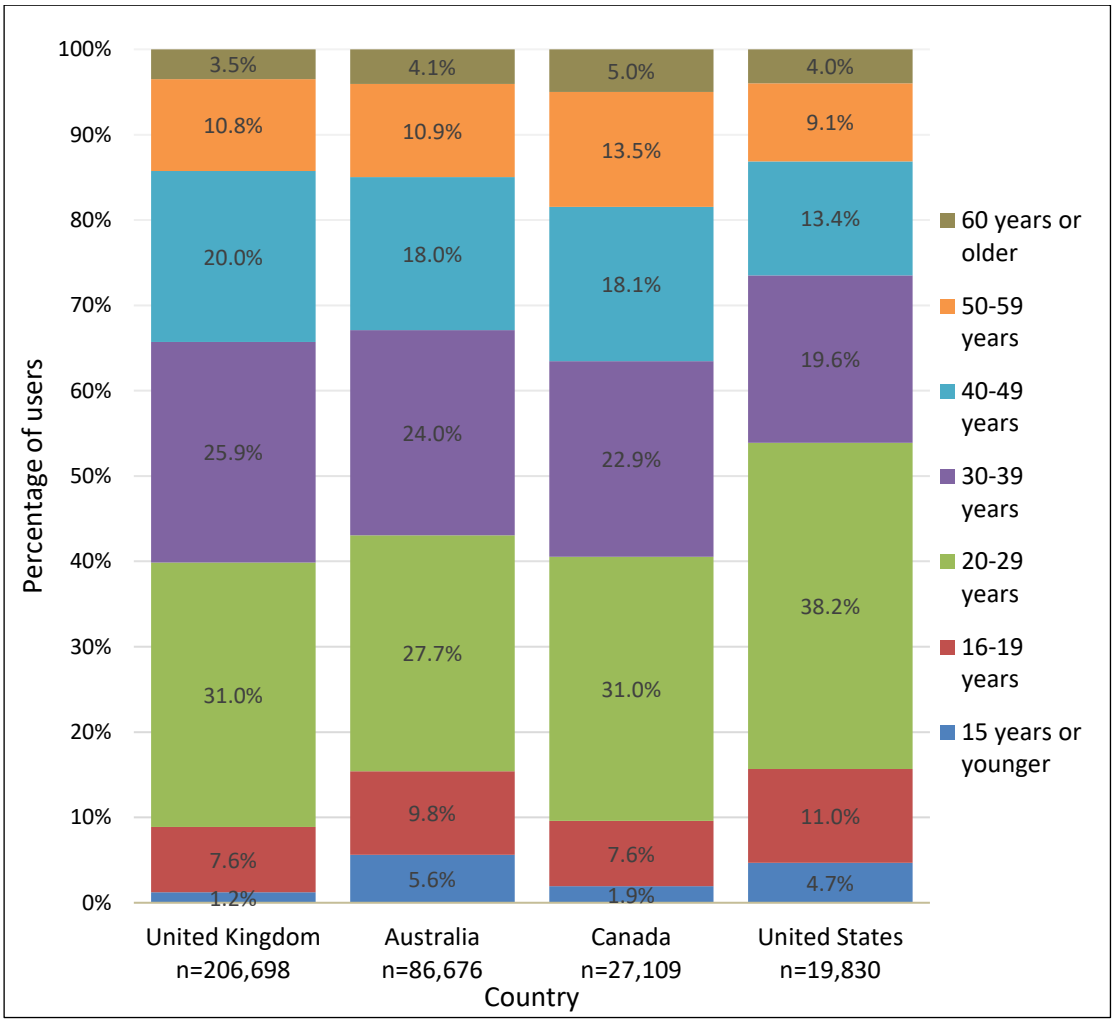

Figure 4.11. Age groups of MoodGYM users by country (n=340,313).

Chi-square tests for comparisons between each country pair were undertaken, using Bonferroni corrections to account for the six multiple comparisons. Significant differences at $p < .001$ were found for all comparisons, shown in Table 4.7. The strongest difference in age group distributions was between users in Canada compared to the United States (Cramer's $V = .148$) with Canadian users more likely to be older, and between users in the United Kingdom and Australia with a larger percentage of Australian users aged under 19 (Cramer's $V = .136$).

Table 4.7. Chi-square test statistics: Age group by pairwise country comparisons.

Comparison	**$\chi^2(6)$**	***n***	***p***	**Bonferroni adjusted *p***	**Cramer's *V***
U.K. vs Australia	5,458.02	293,374	<.001	<.001	.136
U.K. vs Canada	523.48	233,807	<.001	<.001	.047
U.K. vs U.S.	2,687.47	226,528	<.001	<.001	.109
Australia vs Canada	941.52	113,785	<.001	<.001	.091
Australia vs U.S.	1,040.07	106,506	<.001	<.001	.099
Canada vs U.S.	1,022.52	46,939	<.001	<.001	.148

Gender by age group

Figure 4.12 displays the age group distributions separately for male and female users from each country, and Figure 4.13 depicts the percentage of male users within each age group, for users from each country. The overall trend towards older male than female users reported for the sample as a whole (see above) was also observed for users within each country.

Figure 4.12. Age groups by gender of MoodGYM users, by country (n=340,313).

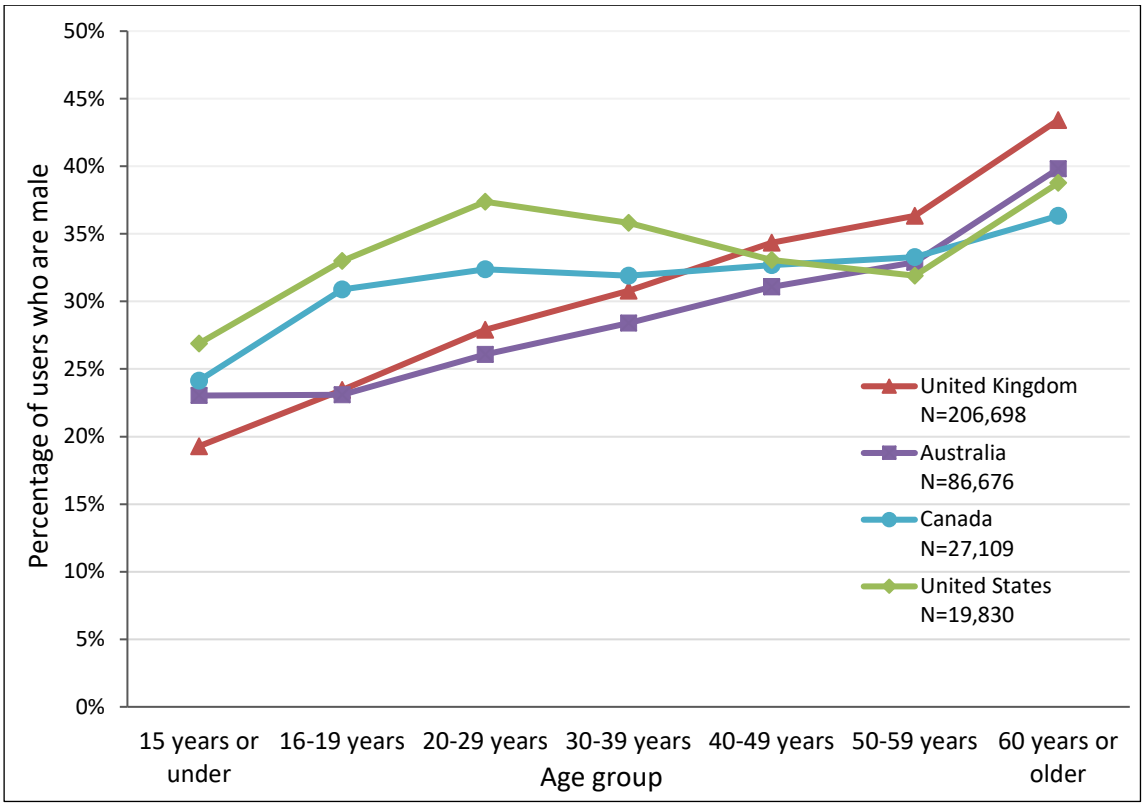

Figure 4.13. Percentage of MoodGYM users within each age group who are male, by country (n=340,313).

Location

The number of users who self-reported being in a rural or remote region was also examined by country of registration. Figure 4.14 displays the percentage of users from rural areas for each country. A Chi-square test showed a statistically significant overall association between rurality and country, χ^2 (3, *n*=340,313) = 5,005.42 *p* < .001. This relationship was weak, Cramer's *V* = .121.

Table 4.8 summarises the findings from six Chi-square tests comparing country pairs, with Bonferroni corrections. All associations were significant. The strongest relationship was between the U.K. and Canada, which have the largest (30.3%) and lowest (16.3%) percentages of users from rural areas, respectively.

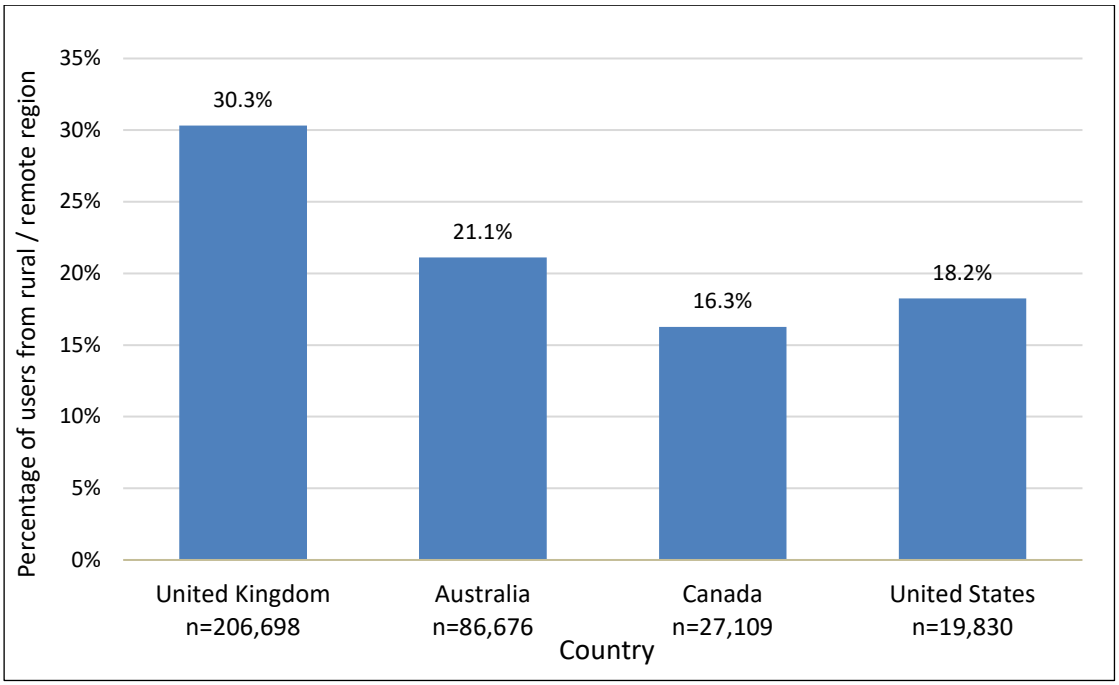

Figure 4.14. Percentage of MoodGYM users from rural/remote region, by country (n=340,313).

Table 4.8. Chi-square test statistics: Rural or not, by pairwise country comparisons.

Comparison	**$\chi^2(6)$**	***n***	***p***	**Bonferroni adjusted *p***	**Phi**
U.K. vs Australia	2,588.81	293,374	<.001	<.001	.094
U.K. vs Canada	2,308.06	233,807	<.001	<.001	.099
U.K. vs U.S.	1,272.43	226,528	<.001	<.001	.075
Australia vs Canada	302.22	113,785	<.001	<.001	.052
Australia vs U.S.	89.84	106,506	<.001	<.001	.028
Canada vs U.S.	31.46	46,939	<.001	<.001	-.026

Education level

The education level of users was also compared between countries (see Figure 4.15). Nearly half of users from both the United Kingdom and Australia reported having either a Bachelor's degree (24.5% U.K.; 25.2% Australia) or a trade, apprenticeship or certificate (26.3% U.K.; 22.1% Australia). Bachelor's graduates also formed the largest percentage of users within Canadian users (24.9%) and users from the U.S. (27.0%).

Overall there was a statistically significant association between education level and country of registration, χ^2 (10, n = 340,313) = 17,086.34, p < .001; however, this relationship was not strong, Cramer's V = .129.

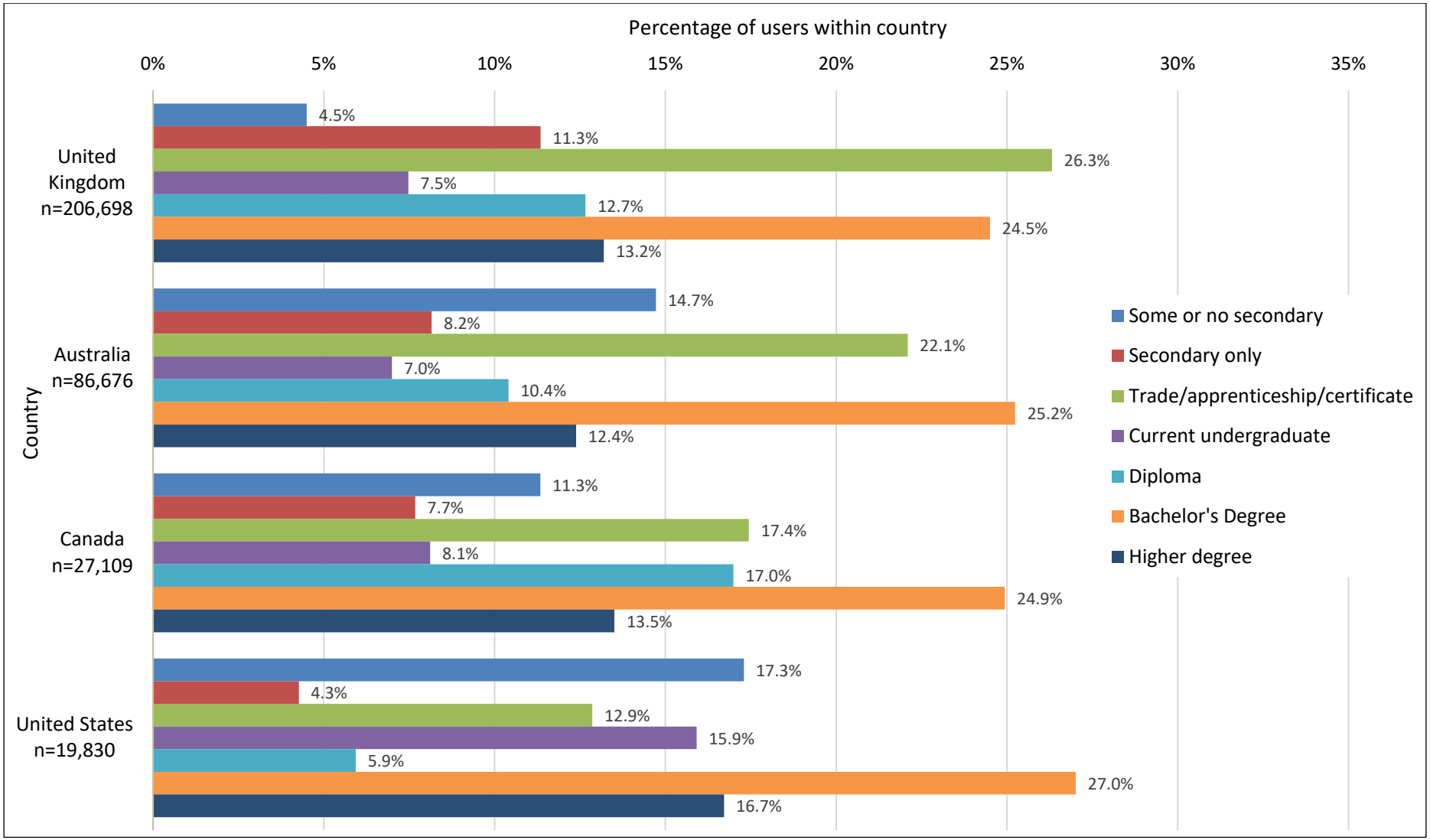

Figure 4.15. Education level of MoodGYM users, by country (n=340,313).

The differences in education level of users were compared between each pair of countries, using multiple Chi-square tests, with Bonferroni corrections for the six comparisons. Significant differences at $p < .001$ were found for all of the pairwise Chi-square tests, summarised at Table 4.9. The biggest differences were between the educational level of users from the U.S. compared to Canada (Cramer's $V = .231$), with larger percentages of U.S. users being current undergraduate students, having a Bachelor's or higher degree, or having only some secondary schooling, and smaller percentages having diplomas or only secondary education. The next biggest difference was between the U.S. and U.K. users, with a much larger proportion of U.K. users having a trade, apprenticeship or certificate.

Table 4.9. Chi-square tests: Education level by pairwise country comparisons.

Comparison	**$\chi^2(6)$**	***n***	***p***	**Bonferroni adjusted *p***	**Cramer's *V***
U.K. vs Australia	9,839.85	293,374	<.001	<.001	.183
U.K. vs Canada	3,517.37	233,807	<.001	<.001	.123
U.K. vs U.S.	9,885.07	226,528	<.001	<.001	.209
Australia vs Canada	1,193.68	113,785	<.001	<.001	.102
Australia vs U.S.	3,145.78	106,506	<.001	<.001	.172
Canada vs U.S.	2,509.36	46,939	<.001	<.001	.231

Pathway to MoodGYM

The method by which users found the *MoodGYM* program was also compared for the four countries with the highest registrations ($n = 338,668$). The 1,645 users who did not respond to this question were not included in the analysis. Figure 4.16 shows different patterns for users of different countries. Visual inspection suggests that the highest source of referral was either by a general practitioner (U.K., Canada) or web search/link (Australia, U.S.). The lowest referral for all countries was by teacher/colleague or media story.

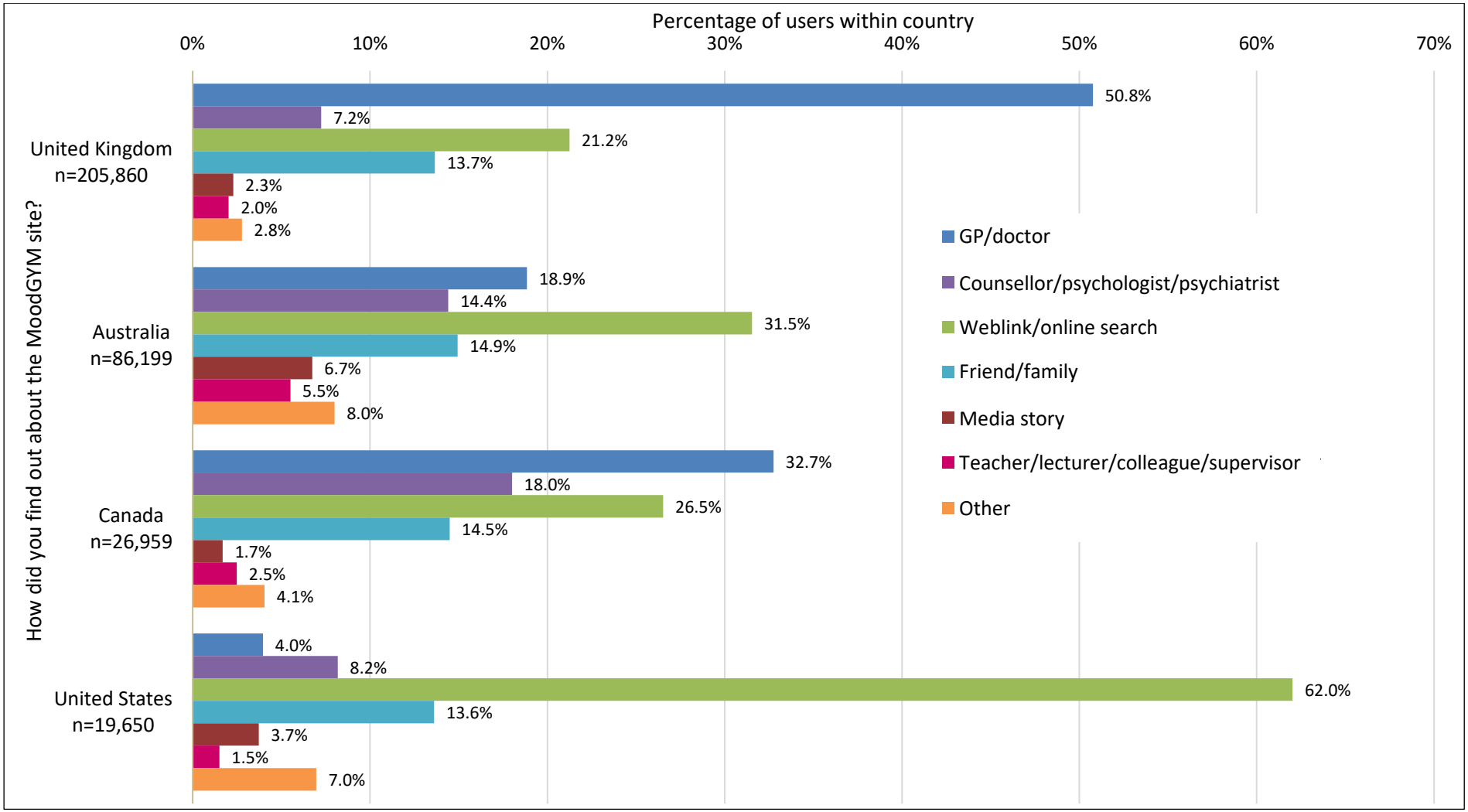

Figure 4.16. Percentage of users who found MoodGYM via different pathways, by country (n=338,668).

Overall the relationship between country and the pathway to the program was statistically significant, χ^2 (18, *n* = 338,668) = 50,865.39, *p* <.001. The level of association between variables was moderate (Cramer's *V* = .224).

Follow-up tests were conducted by examining the adjusted standardised residuals for the contingency table of country by pathway frequencies, which are presented in Table 4.10. A Bonferroni adjustment was made to the *p* values to account for the 28 adjusted residuals to be examined.

For both U.K. and Australia, there were significant differences between the observed frequencies of each of the pathways to the program and the frequencies that would have been expected if country did not influence pathway to the program. For Canada and the U.S., all differences were significant, except for the association between country and being referred by a friend or family member, and in the case of Canada finding the program via online search or a web-link.

The largest influence of country on pathway was between United Kingdom and referral by a doctor (adj. residual = 182.74, *p* < .001), with more U.K. users (50.8%) referred to the program than expected, and between Australia and referral by a doctor (adj. residual = -137.27, *p* < .001), with fewer Australians (18.9%) referred than expected. The next largest influence was that more U.S. users found the program via a web-link or online search (62.0%, adj. residual = 115.54, *p* < .001) than would otherwise be expected.

Table 4.10. Contingency table and adjusted residuals for Chi-square tests between each country and method of finding the program.

	GP / doctor	Counsellor/ psych	Weblink / search	Friend / family	Media Story	Teacher/ colleague	Other
United Kingdom							
n (% U.K.)	104, 506 (50.8%)	14,921 (7.2%)	43,732 (21.2%)	28,112 (13.7%)	4,701 (2.3%)	4,170 (2.0%)	5,718 (2.8%)
Adjusted residual	182.74	-66.09	-88.60	-8.16	-46.47	-38.43	-58.84
p	<.001	<.001	<.001	<.001	<.001	<.001	<.001
Bonferroni adjusted *p*	<.001	<.001	<.001	<.001	<.001	<.001	<.001
Australia							
n (% Australia)	16,250 (18.9%)	12,424 (14.4%)	27,186 (31.5%)	12,880 (14.9%)	5,812 (6.7%)	4,748 (5.5%)	6,899 (8.0%)
Adjusted residual	-137.27	50.26	37.60	8.76	61.21	52.32	58.54
p	<.001	<.001	<.001	<.001	<.001	<.001	<.001
Bonferroni adjusted *p*	<.001	<.001	<.001	<.001	<.001	<.001	<.001
Canada							
n (% Canada)	8,829 (32.7%)	4,855 (18.0%)	7,151 (26.5%)	3,906 (14.5%)	456 (1.7%)	669 (2.5%)	1,093 (4.1%)
Adjusted residual	-20.20	45.82	-0.48	2.17	-16.53	-4.44	-3.31
p	<.001	<.001	.629	.030	<.001	<.001	.001
Bonferroni adjusted *p*	<.001	<.001	---	.833	<.001	<.001	0.026
United States							
n (% U.S.)	780 (4.0%)	1,608 (8.2%)	12,187 (62.0%)	2,676 (13.6%)	732 (3.7%)	297 (1.5%)	1,370 (7.0%)
Adjusted residual	-102.48	-8.67	115.54	-1.78	2.14	-12.07	17.64
p	<.001	<.001	<.001	.075	.033	<.001	<.001
Bonferroni adjusted *p*	<.001	<.001	<.001	---	0.9140	<.001	<.001

Note: A p < .001 indicates the observed value was significantly different from the expected value if the country did not impact on pathway.

Motivation level

The percentage of users from each of the top four registration countries who self-reported to be motivated to complete the program is displayed in Figure 4.17 (n = 340,313). The relationship between the country and number of motivated users was statistically significant, χ^2 (3, n = 340,313) = 113.11 p < .001), but weak, Cramer's V = 0.018.

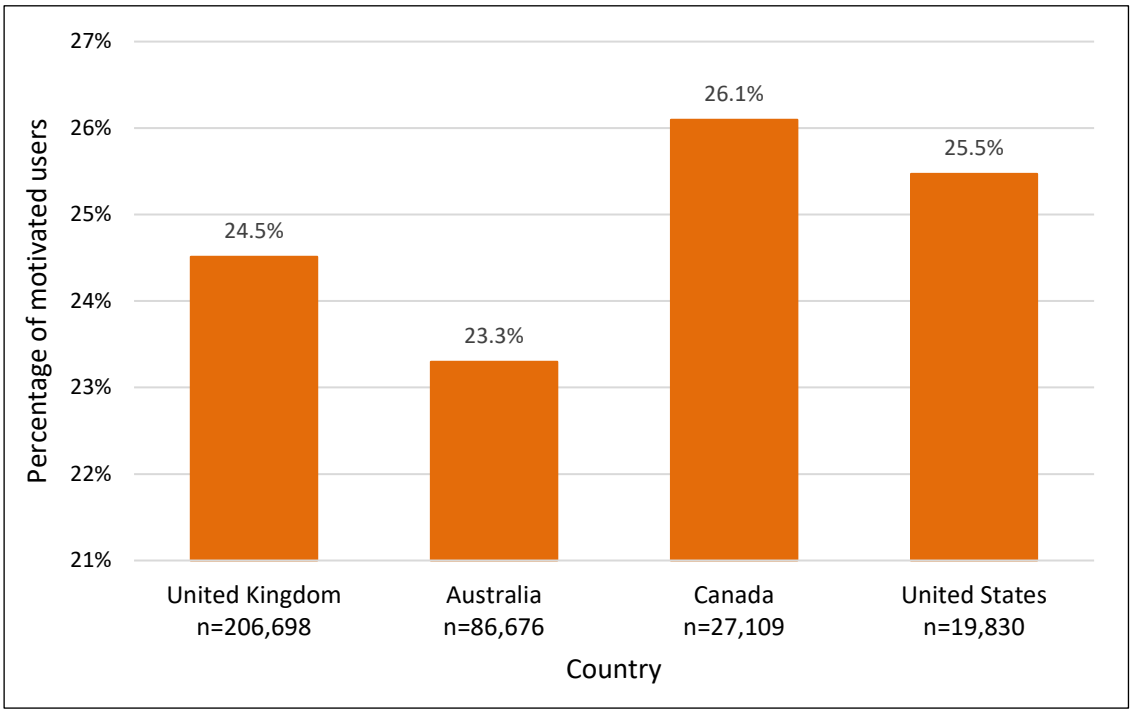

Figure 4.17. Percentage of 'motivated' MoodGYM users, by country (n=340,313).
In answer to the Motivation question, these users answered, "I am committed to working through the entire MoodGYM program".

The differences in percentages of motivated users for each pair of countries was also investigated using Chi-square tests with Bonferroni corrections to account for the six comparisons. There were statistically significant relationships between motivation and country for each pair, except for between the U.S. and the U.K., and between the U.S. and Canada (see Table 4.11). The percentage of motivated users from Canada was significantly higher than for the U.K. and Australia, and the percentage of Australian users who were motivated was significantly less than for each of the other three countries.

Table 4.11. Chi-square test statistics: Motivation by pairwise country comparisons.

Comparison	**$\chi^2(1)$**	***n***	***p***	**Bonferroni adjusted *p***	**Cramer's *V***
U.K. vs Australia	49.41	293,374	<.001	<.001	0.013
U.K. vs Canada	32.19	233,807	<.001	<.001	0.012
U.K. vs U.S.	8.98	226,528	.003	.018	0.006
Australia vs Canada	88.70	113,785	<.001	<.001	0.028
Australia vs U.S.	42.19	106,506	<.001	<.001	0.020
Canada vs U.S.	2.32	46,939	.128	.768	0.007

4.3.3 Aim 3: Referral by a health professional

The demographic characteristics and motivation status of users from the four highest usage countries are displayed in Table 4.12, grouped by health referral status.

Table 4.12. Whether users were referred by a health professional, by demographic characteristics and motivation level (n=340,313).

		Referred by health professional (%)	Not referred (%)	Total
Gender	Female	111,083 (49.1%)	124,956 (52.9%)	236,039
	Male	53,090 (50.9%)	51,184 (49.1%)	104,274
Age group	19 years or younger	18,566 (49.7%)	18,806 (50.3%)	37,372
	20 – 29 years	48,840 (47.0%)	55,181 (53.0%)	104,021
	30 – 39 years	40,162 (47.6%)	44,243 (52.4%)	84,405
	40 – 49 years	31,785 (49.3%)	32,700 (50.7%)	64,485
	50 years or older	24,820 (49.6%)	25,210 (50.4%)	50,030
Education level	Some secondary	13,700 (48.0%)	14,863 (52.0%)	28,563
	Secondary only	19,452 (58.2%)	13,998 (41.8%)	33,450
	Trade /certificate	46,269 (57.2%)	34,565 (42.8%)	80,834
	Undergraduate	10,958 (40.8%)	15,904 (59.2%)	26,862
	Diploma	21,538 (52.6%)	19,431 (47.4%)	40,969
	Bachelor's degree	35,670 (42.1%)	48,979 (57.9%)	84,649
	Higher degree	16,586 (36.9%)	28,400 (63.1%)	44,986
Country	United Kingdom	119,427 (57.8%)	87,271 (42.2%)	206,698
	Canada	13,684 (50.5%)	13,425 (49.5%)	27,109
	United States	2,388 (12.0%)	17,442 (88.0%)	19,830
	Australia	28,674 (33.1%)	58,002 (66.9%)	86,676
Location	Rural / remote	48,840 (54.9%)	40,131 (45.1%)	88,971
	Urban	115,333 (45.9%)	136,009 (54.1%)	251,342
Motivated	Yes	40,091 (48.3%)	42,893 (51.7%)	82,984
	No	124,082 (48.2%)	133,247 (51.8%)	257,329
	Total	**164,173 (48.2%)**	**176,140 (51.8%)**	**340,313**

Table 4.13 shows the results of the logistic regression analysis investigating the association between user demographic status and motivation level (independent variables) and health referral status. The independent variables were added in two blocks, with age group, gender and education level added at Step 1 (Model 1), and country, location and motivation status included at Step 2 (Model 2). Both models were statistically significant (see Chi-square test results, Table 4.13), and the second step resulted in a significant improvement over the previous model (Step χ^2 (4) = 27,112.43, $p < .001$).

The Nagelkerke pseudo R^2 showed that the first model (gender, age group, education level) accounted for only 3.4% of the variance in referral status. Adding country and location to the model accounted for an additional 10.0% of the variance. The Hosmer and Lemeshow test statistic which measures goodness of fit was significant ($p < .001$) for both models, indicating that there was a difference between the values predicted by the model and the observed values. Although usually an indicator of poor model fit, arguably this discrepancy is of less concern in the context of the large size of the current dataset (n = 340,313) for which it would be extremely difficult to build a model that did not display a significant difference [142].

Prediction success rates for the different models are shown in Table 4.14. Adding gender, age and education level alone increased the prediction success rate to 57.4% for the cases used to build the model compared to 51.8% for the constant only model with no independent variables. The final model had a moderate prediction success rate, with 62.8% of cases successfully categorised.

Table 4.13. Logistic regression model test statistics for referral by a health professional.

	Chi-square (*df*, *p*)	**-2 Log likelihood**	**Cox & Snell R^2**	**Nagelkerke R^2**
Model 1 (gender, age, education)	139.58 (8, <.001)	462,479.52	.026	.034
Model 2 (gender, age, education, country, location, motivated)	35,986.00 (15, <.001)	435,367.086	.100	.134

Table 4.14. Prediction success rates for logistic regression models (dependent variable: referral by a health professional).

	Success rate (Referred)	**Success rate (Not referred)**	**Success rate (Overall)**
Constant only model	0%	100%	51.8%
Model 1 (gender, age, education)	55.7%	59.1%	57.4%
Model 2 (gender, age, education, country, location, motivated)	61.3%	64.2%	62.8%

Table 4.15 presents the odds ratios, with 95% confidence intervals, for the independent variables within each model. The Wald statistics and significance of the predictive contribution of each independent variable are also included.

Table 4.15. Hierarchical logistic regression results for predicting whether a user was referred by a health professional (n=340,313).

	Model 1			Model 2		
	Odds ratio (95% CI)	***p***	**Wald**	**Odds ratio (95% CI)**	***p***	**Wald**
Gender						
Female	**0.86***** (0.85 – 0.88)	<.001	371.90	**0.85***** (0.83 – 0.86)	<.001	440.57
Age group ** reference group: ≥ 50 years*						
≤ 19 years	**0.96*** (0.93 – 0.99)	.013	6.17	1.03 (0.99 – 1.06)	.107	2.59
20 - 29 years	0.99 (0.97 – 1.01)	.358	0.84	1.01 (0.99 – 1.03)	.489	0.48
30 - 39 years	0.93 (0.97 – 1.02)	.710	0.14	0.98 (0.96 – 1.01)	.149	2.08
40 - 49 years	0.99 (0.97 – 1.02)	.856	0.03	**0.97**** (0.94 – 0.99)	.007	7.36
Education level ** reference group: Higher degree*						
Some secondary	**1.62***** (1.56 – 1.67)	<.001	812.08	**2.23***** (2.15 – 2.31)	<.001	2,015.20
Secondary only	**2.40***** (2.33 – 2.47)	<.001	3,375.61	**2.17***** (2.10 – 2.24)	<.001	2,467.15
Trade/ certificate	**2.30***** (2.24 – 2.36)	<.001	4,683.86	**2.17***** (2.12 – 2.23)	<.001	3,759.46
Undergraduate	**1.20***** (1.26 – 1.24)	<.001	125.45	**1.27***** (1.23 – 1.31)	<.001	192.53
Diploma	**1.91***** (1.86 – 1.97)	<.001	2,170.95	**1.78***** (1.73 – 1.83)	<.001	1,600.62
Degree	**1.25***** (1.22 – 1.28)	<.001	346.39	**2.82***** (2.77 – 2.87)	<.001	349.56
Country ** reference group: Australia*						
United Kingdom				**2.82***** (2.77 – 2.87)	<.001	13,976.79
Canada				**2.17***** (2.11-2.23)	<.001	2,890.05
United States				**0.29***** (0.28 – 0.30)	<.001	2,830.74
Location						
Rural / remote				**1.21***** (1.19 – 1.23)	<.001	529.26
Motivated						
Motivated				1.01 (1.00-1.03)	.106	2.61

* $p \leq .05$ **$p \leq .01$ ***$p < 0.001$

Gender and all education levels were significant predictors in Model 1. Being aged 19 years or younger (compared to aged 50 years or older) was also a significant predictor (p = .013) although this contribution was very small compared to other variables (Wald = 6.17). Gender and education levels remained statistically significant with the addition of country and location in Model 2, and location and all country categories were also highly significant. Whether a user reported that they were motivated to complete the program was not a significant predictor of health referral status. After controlling for country and location, age group was not a significant predictor except for the 40-49 years age group (p = .007). However, the contribution of this age group was small (Wald = 7.20).

Within the final model, controlling for other demographic variables, the odds that health professionals were the source of referral was 15.3% less for female compared to male users. Compared to Australian registrants, users from the United Kingdom and Canada were 2.82 times and 2.17 times more likely respectively to be referred by a health professional. Users from the United States were 70.9% less likely to be referred than Australian users. Further, users from rural and remote regions were 21.2% more likely to be referred by a health professional than those from urban areas. Finally, education level was a significant predictor of referral, with less educated users being more likely to have found the program via a health professional. In particular, compared to users with a higher degree, users who had some secondary education, secondary only, or a trade/certificate were over two times more likely to have been referred.

4.3.4 Aim 4: Initial depression and anxiety symptom scores

Of the 417,354 eligible participants, 372,442 (89.2%) completed the pre-program Goldberg Depression scale, and 386,981 (88.4%) completed the initial Goldberg Anxiety scale. The mean depression score for these users was 6.29 (n = 372,442, SD = 1.88) and the mean anxiety score was 6.42 (n = 386,981, SD = 1.98).

Sub-group comparisons: univariate analyses

Gender

The frequency distributions of depression and anxiety scores for male and female users are shown in Figure 4.18 and Figure 4.19. The mean depression score was 6.28 (n = 117,783, SD=1.92) for male users, and 6.29 (n = 254,659, SD = 1.86) for female users. The mean anxiety scores were 6.21 (n = 116,506, SD = 2.04) and 6.52 (n = 252,475, SD = 1.94), for male and female users respectively. A Welch's t-test showed that the difference in mean Goldberg Depression scores for male and female users was not statistically significant, t(222,885.46) = -0.454, p = .65 (mean difference = 0.003, 95%

CI: -0.010, 0.016). However, the mean Goldberg Anxiety score was significantly higher for female than male users, $t(217{,}038{,}64) = -43.81$, $p < .001$, $d = 0.15$ (mean difference = 0.31, 95% CI: 0.30, 0.32).

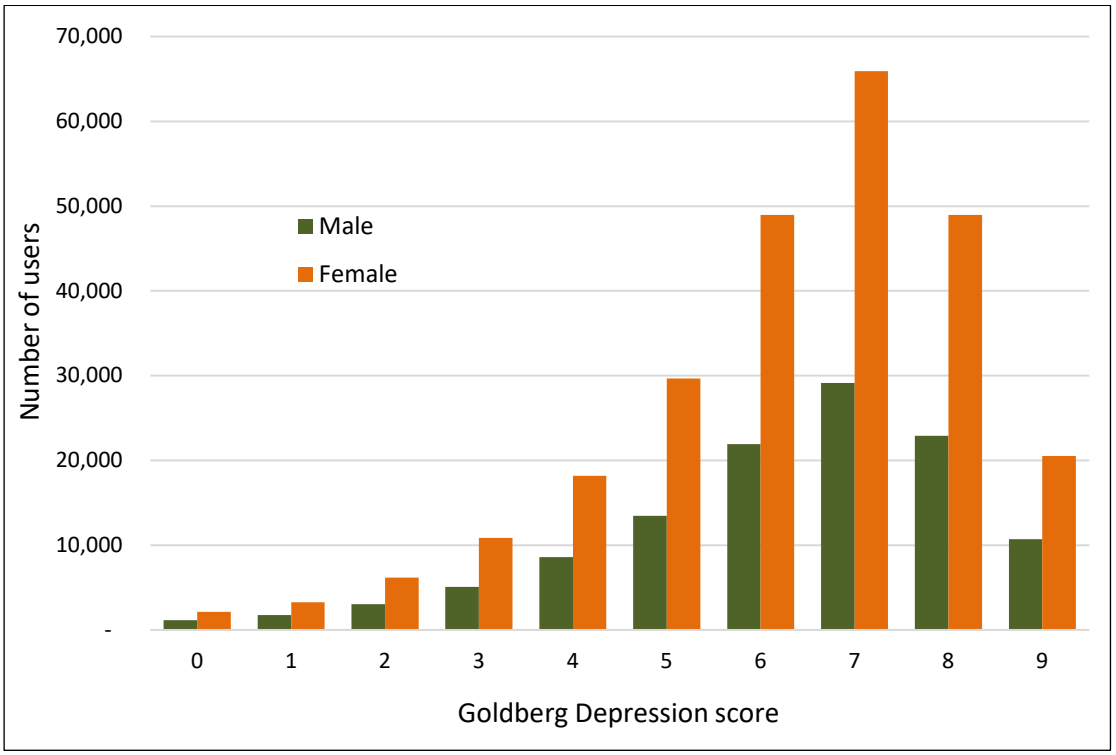

Figure 4.18. Initial depression scores for male and female users (n=372,442).

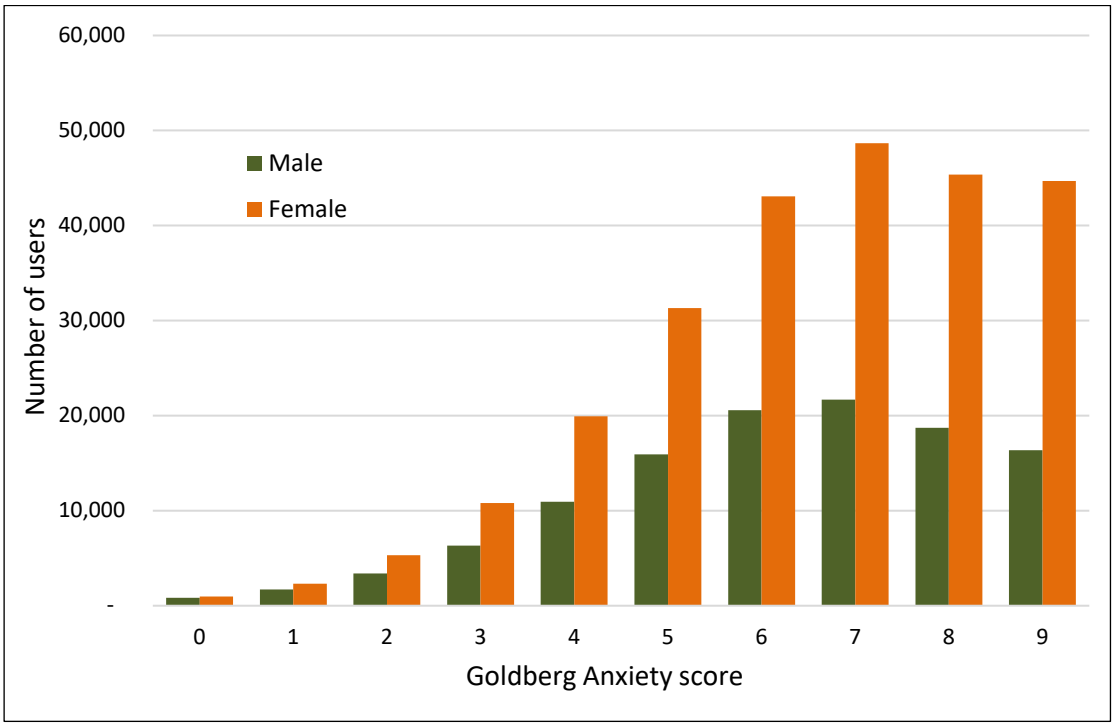

Figure 4.19. Initial anxiety scores for male and female users (n=368,981).

Age group

Figure 4.20 shows the mean initial depression and anxiety scores for users, by age group (error bars represent one standard deviation). Levene's test confirmed that the assumption of homogeneity of variance across age groups had been violated for both the depression and anxiety variables. A one-way between-subjects ANOVA showed a significant effect of age group on initial depression score, Welch's $F(4, 148{,}491.86) = 49.15$, $p < .001$, and initial anxiety score, Welch's $F(4, 146{,}647.37) = 700.50$, $p < .001$. The effect sizes η^2 were .0005 and .0078, for depression and anxiety, respectively, indicating that age accounted for less than 1% of the variance in initial depression and anxiety scores.

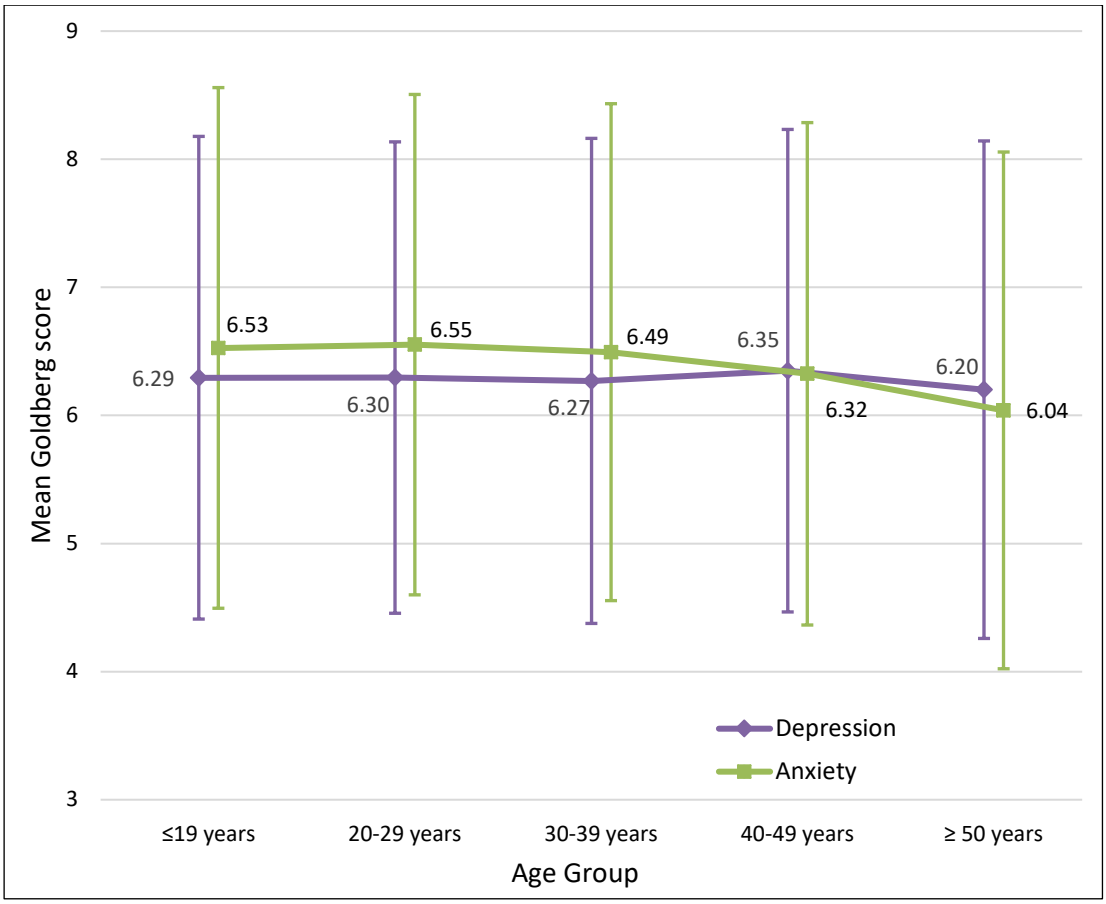

Figure 4.20. Mean initial depression and anxiety scores (+/- SD), by age group (n=372,442 depression; n=368,981 anxiety).

Games-Howell post hoc tests revealed significant differences between all age groups, except for comparisons between the younger categories (see Table 4.16). The differences between users aged 19 years or younger and those aged 20-29 years were not significant for either depression or anxiety, and the difference between the 19 and under group and the 30-39 years group was not significant for depression, and extremely small for anxiety. The largest difference for mean initial depression

scores was between users younger than 19 (higher scores) and those older than 50 (mean difference = 0.148 [95% CI: 0.12, 0.18], $p < .001$). For initial anxiety scores, the largest differences were between the 50 years and older group (lowest scores) and the younger age groups: the mean difference between the 20-29 age group and the 50 plus group was 0.513 ([95% CI: 0.48, 0.54], $p < .001$); it was 0.488 between the youngest and older age groups ([95% CI: 0.45, 0.52], $p < .001$).

Table 4.16. Games-Howell post hoc analysis of initial depression and anxiety scores – comparisons between age groups.

	Depression		**Anxiety**	
	Mean difference (95% CI)	***p***	**Mean difference (95% CI)**	***p***
< 19 years compared to:				
20-29 years	- 0.001 (-0.03, 0.03)	1.00	- 0.026 (-0.06, 0.01)	.183
30-39 years	0.025 (-0.01, 0.06)	.183	**0.033*** (0.00, 0.07)	.049
40-49 years	**- 0.055***** (-0.09, -0.02)	<.001	**0.202***** (0.17, 0.24)	<.001
> 50 years	**0.093***** (0.06, 0.13)	<.001	**0.488***** (0.45, 0.52)	<.001
20-29 years compared to:				
30-39 years	**0.026*** (0.00, 0.05)	.013	**0.059***** (0.04, 0.08)	<.001
40-49 years	**- 0.054***** (-0.08, -0.03)	<.001	**0.228***** (0.20, 0.25)	<.001
> 50 years	**0.095***** (0.07, 0.12)	<.001	**0.513***** (0.48, 0.54)	<.001
30-39 years compared to:				
40-49 years	**-0.080***** (-0.11, 0.05)	<.001	**0.169***** (0.14, 0.20)	<.001
> 50 years	**0.069***** (0.04,0.10)	<.001	**0.455***** (0.42,0.48)	<.001
40-49 years compared to:				
> 50 years	**0.148***** (0.12,0.18)	<.001	**0.286***** (0.25,0.32)	<.001

*** $p < .001$

Gender and age group

Figure 4.21 and Figure 4.22 show the mean initial depression and anxiety scores for male and female users by age group. Error bars represent one standard deviation. A one-way ANOVA was conducted to investigate the effect of age by gender (ten categories) on initial symptom measures, and showed

a significant effect for both depression, Welch's $F(9, 102{,}950.63) = 97.13$, $p < .001$, $\eta^2 = .0025$, and anxiety, Welch's $F(9, 101{,}522.93) = 635.07$, $p < .001$, $\eta^2 = .0158$.

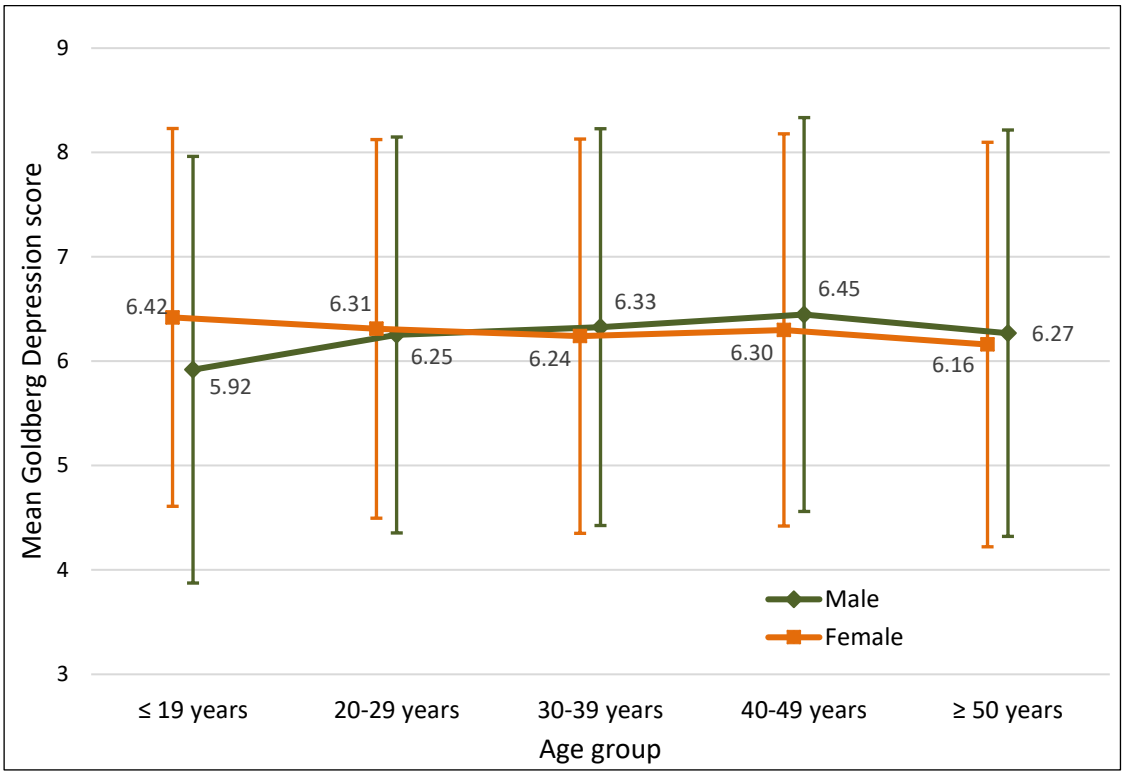

Figure 4.21. Mean initial depression score (+/- SD), by age group and gender (n=372,442).

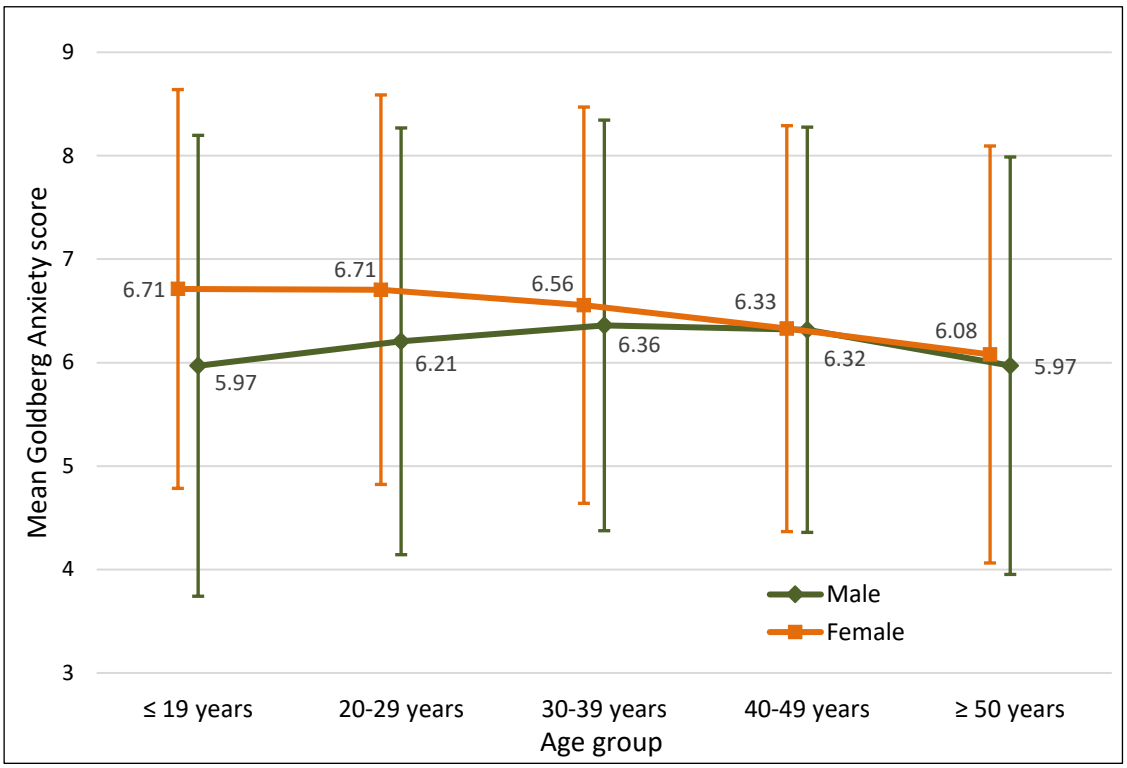

Figure 4.22. Mean initial anxiety score (+/- SD), by age group and gender (n=368,981).

Games-Howell post hoc tests were undertaken, and comparisons between male and female users within each age group were examined (see Table 4.17). There were significant differences in depression scores by gender within each age group. The largest difference was between female and male users younger than 19 (mean difference = 0.501 [95% CI: 0.43, 0.57], $p < .001$), with young female users having higher initial depression scores. In contrast, female users in the 40-49 and 50 plus age groups had significantly lower initial depression scores than male users in the same age ranges.

There were also significant differences in mean anxiety scores for male and female users within all age groups, except for 40-49 years. Within the other four age groups, female users had significantly higher anxiety scores, with the largest differences being for the youngest age groups: the mean difference by gender was 0.743 ([95% CI: 0.66, 0.82], $p < .001$) for the ≤ 19 years group and 0.499 for the 20-29 years group ([95% CI: 0.45, 0.52], $p < .001$).

Table 4.17. Games-Howell post hoc analysis of initial depression and anxiety scores for age by gender sub-groups – comparisons between male and female users for each age group.

	Depression		**Anxiety**	
	Mean difference (95% CI)	***p***	**Mean difference (95% CI)**	***p***
Female compared to male users:				
≤ 19 years	**0.501***** (0.43, 0.57)	<.001	**0.743***** (0.66, 0.82)	<.001
20-29 years	**0.063***** (0.33, 0.46)	<.001	**0.499***** (0.46, 0.54)	<.001
30-39 years	**- 0.084***** (-0.13, -0.04)	<.001	**0.196***** (0.15, 0.24)	<.001
40-49 years	**- 0.148***** (-0.29, -0.10)	<.001	0.011 (-0.04, 0.06)	1.00
≥ 50 years	**-0.107***** (-0.16, -0.05)	<.001	**0.108***** (0.05, 0.17)	<.001

$^{***}p < .001$

Country of registration

The mean initial depression and anxiety scores for users from each country are displayed in Figure 4.23. One-way between-subjects ANOVAs were performed to compare the symptom scores from different countries. Since Levene's test confirmed unequal variances, Welch's *F* values are reported. The analyses showed a significant effect of country on initial depression score, Welch's $F(3, 51{,}910.96) = 609.41$, $p < .001$, and on initial anxiety score, Welch's $F(3, 51{,}036.34) = 407.12$, $p < .001$. The effect sizes η^2 were .006 and .004, for depression and anxiety, respectively.

Users from the United Kingdom had the highest mean Goldberg Depression score (6.45, *n* = 186,234, *SD* = 1.81) and the highest Goldberg Anxiety score (6.58, *n* = 184,732, *SD* = 1.87). Users from the United States had the lowest mean symptom scores (Goldberg Depression: 6.06, *n* = 17,366, *SD* = 1.94; Goldberg Anxiety: 6.17, *n* = 17,215, *SD* = 2.14).

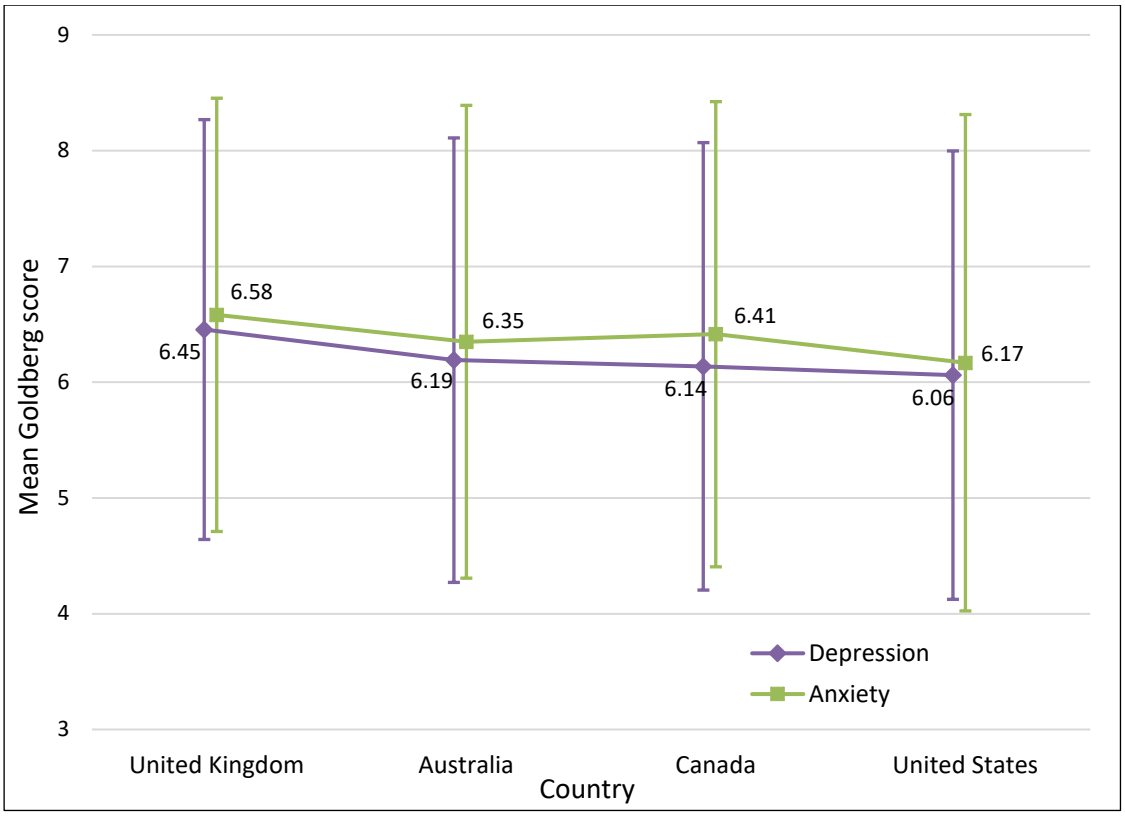

Figure 4.23. Mean initial depression and anxiety scores (+/- SD) for users from top four countries (n=304,041, depression; n=301,374, anxiety).

Games-Howell post hoc tests showed significant differences for all pairwise country comparisons (see Table 4.18). The largest differences for both mean depression and anxiety scores were between users from the United Kingdom (higher scores) compared to those from the United States (mean depression difference = 0.394 [95% CI: 0.35, 0.43], $p < .001$; mean anxiety difference = 0.414 [95% CI: 0.37, 0.46], $p < .001$).

Table 4.18. Games-Howell post hoc analysis of initial depression and anxiety scores – comparisons between countries.

	Depression		**Anxiety**	
	Mean difference (95% CI)	***p***	**Mean difference (95% CI)**	***p***
U.K. compared to:				
Australia	**0.265***** (0.24, 0.29)	<.001	**0.233***** (0.21, 0.25)	<.001
Canada	**0.318***** (0.28, 0.35)	<.001	**0.167***** (0.13, 0.20)	<.001
U.S.	**0.394***** (0.35, 0.43)	<.001	**0.414***** (0.37, 0.46)	<.001
Australia compared to:				
Canada	**0.054**** (0.02, 0.09)	.001	**-0.065***** (-0.10, -0.03)	<.001
U.S.	**0.129***** (0.09, 0.17)	<.001	**0.181***** (0.14, 0.23)	<.001
Canada compared to:				
U.S.	**0.076**** (0.03, 0.13)	.001	**0.246***** (0.19, 0.30)	<.001

** $p < .01$ ***$p < .001$.

Note: A positive mean difference signifies a higher symptom score in the reference country.

Location

The mean Goldberg Depression score for users who self-reported living in a rural or remote region was 6.40 (n = 92,125, SD = 1.84), compared to 6.25 (n = 280,317, SD = 1.90) for users residing in a city. A Welch's t-test showed that the difference was statistically significant, t(161,498.54) = 22.01, $p < .001$, d = 0.084. Rural users also had a higher mean anxiety score (M = 6.54, n = 91,322, SD = 1.92) than non-rural users (M = 6.38, n = 277,659, SD = 1.99). A Welch's t-test showed that this difference was also significant, t(160,526.26) = 20.92, $p < .001$, d = 0.080. Cohen's d effect sizes indicated that both effects were small (both were less than 0.1).

Education level

Figure 4.24 shows the mean Goldberg assessment scores according to users' education level. A one-way between-subjects ANOVA was performed to examine the effect of education level on both the initial depression and the initial anxiety scores. Levene's tests were significant for both analyses, indicating violation of the assumption of homogeneity of variance. The analyses showed a significant effect of education level on initial depression score, Welch's F(6, 132,052.72) = 847.53, $p < .001$, η^2 = .014, and initial anxiety score, Welch's F(6, 130,203.87) = 924.86, $p < .001$, η^2 = .015.

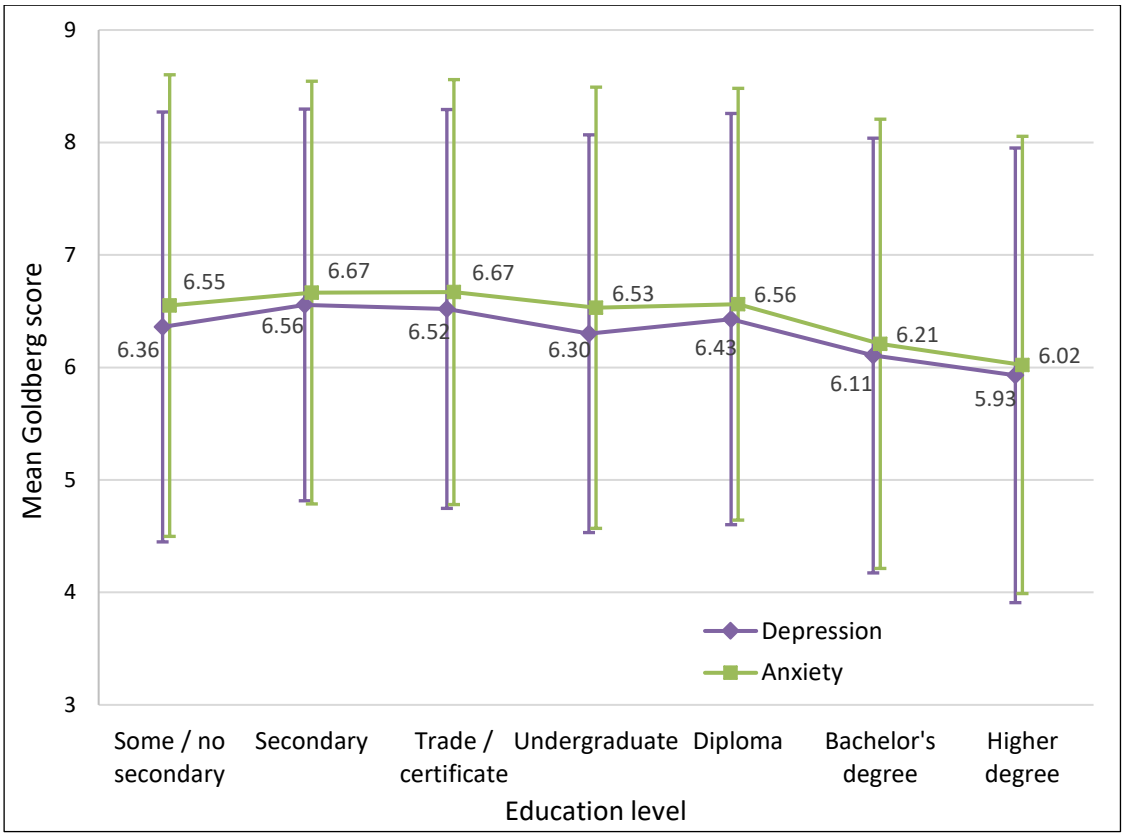

Figure 4.24. Mean initial depression and anxiety scores (+/- SD), by users' education level (n=372,442, depression; n=368,981, anxiety).

Games-Howell post hoc tests revealed significant differences between most groups (see Table 4.19). For mean depression scores, all comparisons were significant, with the largest differences being between users with a higher degree or Bachelor's degree (lower scores) and those with other levels of education. For example, the mean difference between users with secondary education only compared to those with a higher degree was 0.63 (95% CI: 0.59, 0.67), $p < .001$. The absolute mean depression scores were highest for users with secondary education only ($M = 6.56$, $n = 33,498$, $SD = 1.74$) and lowest for users with a higher degree ($M = 5.93$, $n = 55,395$, $SD = 1.88$).

For anxiety, the pairwise comparisons between the least educated users (those who had not completed secondary education), those who were current undergraduates, and those who had achieved a diploma, were all non-significant. The differences between users who had completed secondary education only, and those that had a trade, apprenticeship or certificate, were also not significant. These users with secondary education or a trade, apprenticeship or certificate had the

highest absolute mean anxiety scores (secondary: $M = 6.67$, $n = 33,186$, $SD = 1.88$; trade: $M = 6.67$, $n = 81,018$, $SD = 1.89$), significantly higher than users in other education levels, with the largest difference being with users with a higher degree (secondary compared to higher degree: mean

difference = 0.643 [95% CI: 0.60, 0.68], $p < .001$; trade compared to higher degree: mean difference = 0.647 [95% CI: 0.62, 0.68], $p < .001$). Similar to depression scores, the mean anxiety scores were lowest for users with a higher degree (M = 6.02, n = 54,874, SD = 2.02).

Table 4.19. Games-Howell post hoc analysis of initial depression and anxiety scores – comparisons between education levels.

	Depression		**Anxiety**	
	Mean difference (95% CI)	***p***	**Mean difference (95% CI)**	***p***
Some/no secondary compared to:				
Secondary only	**- 0.200***** (-0.24, -0.16)	<.001	**- 0.119***** (-0.16, -0.07)	<.001
Trade /certificate	**- 0.165***** (-0.20, -0.23)	<.001	**- 0.123***** (-0.16, -0.08)	<.001
Undergraduate	0.053** (0.01, 0.10)	.005	0.016 (-0.03, 0.06)	.953
Diploma	**-0.070***** (-0.11, -0.03)	<.001	-0.015 (-0.06, 0.03)	.946
Bachelor's Degree	**0.250***** (0.21, 0.29)	<.001	**0.335***** (0.30, 0.37)	<.001
Higher degree	**0.431***** (0.39, 0.47)	<.001	**0.524***** (0.48, 0.57)	<.001
Secondary only compared to:				
Trade /certificate	0.035* (0.00, 0.07)	0.036	-0.004 (-0.04, 0.03)	1.000
Undergraduate	**0.253***** (0.21, 0.29)	<.001	**0.135***** (0.09, 0.18)	<.001
Diploma	**0.130***** (0.09, 0.17)	<.001	**0.104***** (0.06, 0.14)	<.001
Bachelor's Degree	**0.450***** (0.42, 0.48)	<.001	**0.454***** (0.42, 0.49)	<.001
Higher degree	**0.630***** (0.59, 0.67)	<.001	**0.643***** (0.60, 0.68)	<.001

Table 4.19. Games-Howell post hoc analysis of initial depression and anxiety scores – comparisons between education levels (continued).

	Depression		Anxiety	
	Mean difference (95% CI)	***p***	**Mean difference (95% CI)**	***p***
Trade / certificate compared to:				
Undergraduate	**0.218***** (0.18, 0.25)	<.001	**0.139***** (0.10, 0.18)	<.001
Diploma	**-0.095***** (0.06, 0.13)	<.001	**0.108***** (0.07, 0.14)	<.001
Bachelor's Degree	**0.415***** (0.39, 0.44)	<.001	**0.458***** (0.43, 0.49)	<.001
Higher degree	**0.596***** (0.56, 0.63)	<.001	**0.647***** (0.62, 0.68)	<.001
Undergraduate compared to:				
Diploma	**- 0.123***** (-0.16, -0.08)	<.001	-0.031 (-0.07, 0.01)	0.314
Bachelor's Degree	**0.197***** (0.16, 0.23)	<.001	**0.319***** (0.28, 0.36)	<.001
Higher degree	**0.377***** (0.34, 0.42)	<.001	**0.508***** (0.47, 0.55)	<.001
Diploma compared to:				
Bachelor's Degree	**0.320***** (0.29, 0.35)	<.001	**0.351***** (0.32, 0.38)	<.001
Higher degree	**0.501***** (0.46, 0.54)	<.001	**0.540***** (0.50, 0.58)	<.001
Bachelor's Degree compared to:				
Higher degree	**0.180***** (0.15, 0.21)	<.001	**0.189***** (0.16, 0.22)	<.001

*p < .05 **p < .01 ***p < .001

Health professional referral

Users who were referred to the program by a health professional (GP or doctor, or counsellor, psychologist or psychiatrist) had higher symptom scores than users who were not. The mean depression score for referred users was 6.51 (n = 159,856, SD = 1.80) compared to 6.12 (n = 159,856, SD = 1.93) for users who found the program via another route. Mean anxiety scores were also higher for referred users (M = 6.69, n = 158,566, SD = 1.84) compared to those who were not referred by a practitioner (M = 6.22, n = 210,415, SD = 2.05). Welch's t-tests showed that these differences were

significant, with small effect sizes [depression: $t(355{,}683.16) = 62.87$, $p < .001$, $d = 0.208$; anxiety: $t(357{,}835.94) = 74.18$, $p < .001$, $d = 0.247$].

History of depression

Initial depression and anxiety scores were compared for users according to whether they reported a history of depression. Those with a history of depression had higher Goldberg Depression scores ($M = 6.52$, $n = 314{,}947$, $SD = 1.73$) than those who answered 'no' to this question ($M = 5.01$, $n = 57{,}495$, $SD = 2.12$). A Welch's t-test showed that the difference was statistically significant, $t(71{,}423.16) = 157.61$, $p < .001$. Cohen's d was 0.715, indicating a medium effect.

Users with depression history also had higher Goldberg Anxiety scores ($M = 6.55$, $n = 312{,}156$, $SD = 1.92$) compared to those without ($M = 5.69$, $n = 56{,}825$, $SD = 2.10$). A Welch's t-test revealed that this difference was also significant; however, the effect size was smaller, $t(75{,}202.78) = 91.650$, $p < .001$, $d = 0.418$.

Treatment history

The mean depression and anxiety scores for users were also examined according to the number of depression treatment options they had been tried previously. Figure 4.25 shows a weak positive linear relationship between number of prior treatments and symptom score (depression: $r(372{,}442) = .14$, $p < .001$; anxiety: $r(368{,}981) = .13$, $p < .001$). Regression analyses showed that number of treatment options was a significant predictor of both initial depression severity ($\beta = .14$, $t(372{,}440) = 85.96$, $p < .001$) and initial anxiety severity ($\beta = .13$, $t(368{,}979) = 80.91$, $p < .001$). However, number of treatments accounted for only 2% of the variance in both the model for depression ($R^2 = .019$, $F(1, 372{,}442) = 7{,}389.68$, $p < .001$) and the model for anxiety ($R^2 = .017$, $F(1, 368{,}979) = 6{,}546.07$, $p < .001$).

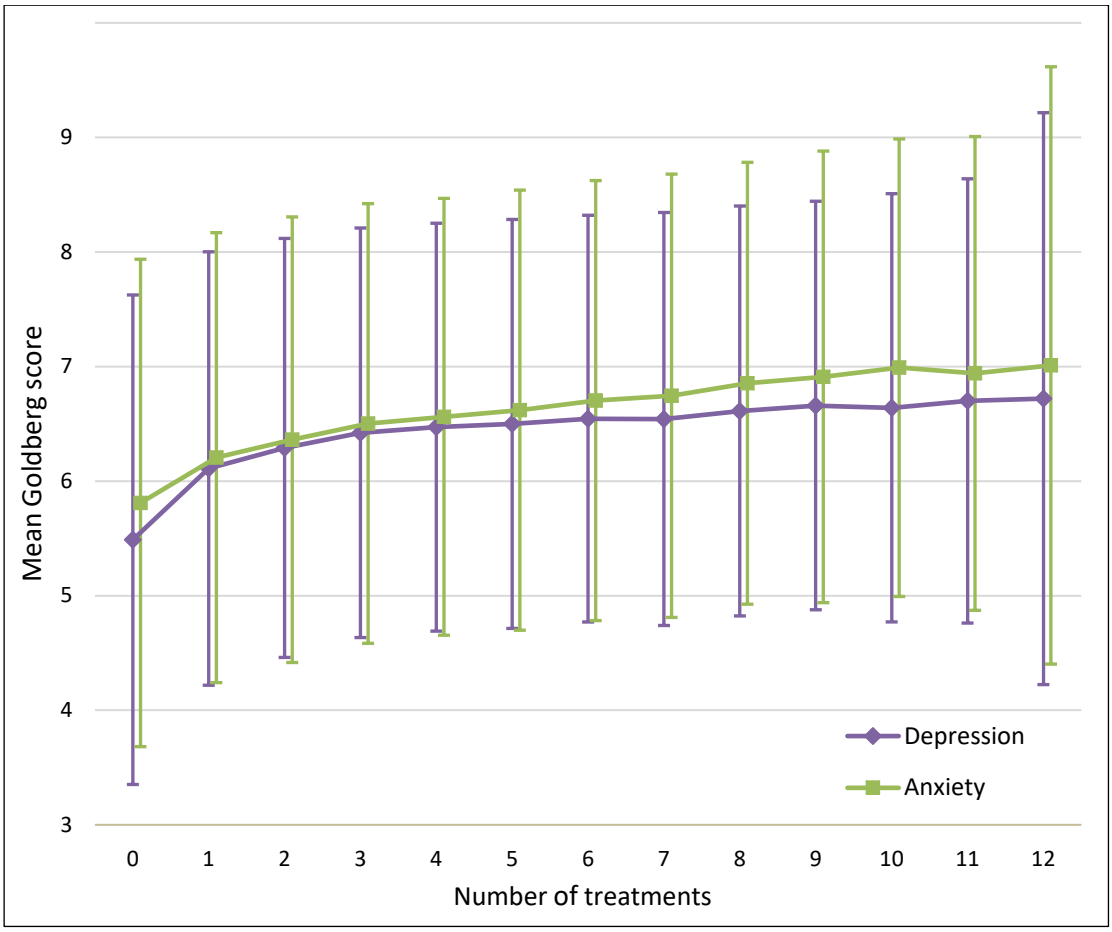

Figure 4.25. Mean initial depression and anxiety scores (+/- SD) as a function of the number of depression treatments previously tried (n=372,442, depression; n=368,981, anxiety).

Motivation level

The mean Goldberg symptom scores for users who indicated that they were committed to working through the entire program were higher than for other users. These 'motivated' users had a mean depression score of 6.38 (n = 93,457, SD = 1.90), compared to 6.25 for other users (n = 278,985, SD = 1.87). The mean anxiety score for motivated users was 6.55 (n = 92,675, SD = 1.96), compared to a mean of 6.38 (n = 276,306, SD = 1.98) for other users. A Welch's t-test showed that these differences were statistically significant, however, the corresponding effect sizes were extremely small [depression: t(158,524.18) = 17.65, p < .001, d = .07; anxiety: t(160,375.67) = 22.71, p < .001, d = .09].

Predicting baseline depression and anxiety scores: Multivariate analyses

Users were included in the multiple regression analysis if they were from one of the four countries with the highest number of users and had completed both the initial Goldberg Depression and Anxiety measures (n = 301,373). The mean depression and anxiety scores for users from each independent variable sub-group are displayed in Table 4.20.

Table 4.20. Mean depression and anxiety scores for users included in the linear regression analyses, by sub-groups (n=301,373).

	n (%)	Goldberg Depression *M (SD)*	Goldberg Anxiety *M (SD)*
Gender			
Female	209,967 (69.7%)	6.34 (1.85)	6.57 (1.92)
Male	91,406 (30.3%)	6.35 (1.89)	6.29 (2.00)
Age group			
≤ 19 years	33,583 (11.1%)	6.33 (1.88)	6.57 (2.02)
20 – 29 years	93,813 (31.1%)	6.35 (1.82)	6.63 (1.92)
30 – 39 years	74,923 (24.9%)	6.32 (1.87)	6.55 (1.91)
40 – 49 years	56,480 (18.7%)	6.41 (1.86)	6.39 (1.93)
≥ 50 years	42,574 (14.1%)	6.28 (1.91)	6.13 (1.98)
Country			
United Kingdom	184,731 (61.3%)	6.46 (1.81)	6.58 (1.87)
Canada	23,501 (7.8%)	6.14 (1.93)	6.41 (2.01)
United States	17,215 (5.7%)	6.06 (1.93)	6.17 (2.14)
Australia	75, 926 (25.2%)	6.19 (1.91)	6.35 (2.04)
Location			
Rural / remote	79,369 (26.3%)	6.44 (1.82)	6.57 (1.90)
Urban	222,004 (73.7%)	6.31 (1.87)	6.46 (1.96)
Education level			
Some secondary	24,870 (8.3%)	6.38 (1.92)	6.57 (2.05)
Secondary only	29,808 (9.9%)	6.58 (1.73)	6.70 (1.86)
Trade /certificate	71,237 (23.6%)	6.55 (1.76)	6.70 (1.87)
Undergraduate	24,290 (8.1%)	6.35 (1.75)	6.60 (1.93)
Diploma	36,020 (12.0%)	6.45 (1.82)	6.60 (1.90)
Bachelor's degree	75,190 (24.9%)	6.15 (1.92)	6.27 (1.97)
Higher degree	39,958 (13.3%)	6.01 (1.99)	6.13 (2.00)
Health professional referral			
Referred	146,111 (48.5%)	6.52 (1.79)	6.71 (1.83)
Not referred	155,262 (51.5%)	6.17 (1.91)	6.27 (2.03)
History of depression			
History	254,096 (84.3%)	6.58 (1.70)	6.62 (1.89)
No history	47,277 (15.7%)	5.08 (2.15)	5.76 (2.08)

Table 4.20. Mean depression and anxiety scores for users included in the linear regression analyses, by sub-groups (n=301,373) (continued).

	n (%)	Goldberg Depression *M (SD)*	Goldberg Anxiety *M (SD)*
Previous treatments			
0	34,666 (11.5%)	5.48 (2.14)	5.86 (2.12)
1	43,483 (14.4%)	6.16 (1.87)	6.27 (1.93)
2	50,992 (16.9%)	6.36 (1.79)	6.43 (1.91)
3	48,894 (16.2%)	6.49 (1.76)	6.57 (1.88)
4	41,088 (13.6%)	6.54 (1.75)	6.63 (1.88)
5	32,199 (10.7%)	6.56 (1.76)	6.68 (1.89)
6	22,862 (7.6%)	6.61 (1.75)	6.77 (1.89)
7	14,311 (4.7%)	6.60 (1.79)	6.81 (1.91)
8	7,620 (2.5%)	6.68 (1.76)	6.93 (1.90)
9	3,563 (1.2%)	6.70 (1.79)	6.99 (1.93)
10	1,319 (0.4%)	6.70 (1.84)	7.06 (1.97)
11	328 (0.1%)	6.72 (1.85)	7.01 (2.08)
12	48 (0.02%)	6.29 (2.64)	7..04 (2.64)
Motivation level			
Motivated	75,588 (24.7%)	6.43 (1.86)	6.61 (1.93)
Not motivated	226,785 (75.3%)	6.31 (1.85)	6.45 (1.95)
Total	**301,373 (100%)**	**6.34 (1.86)**	**6.49 (1.95)**

The multiple regression analyses were conducted with all independent variables entered simultaneously. These variables predicted both the initial depression score, $F(19, 301{,}353) = 2{,}022.29$, $p < .001$, $R^2 = .113$, and the initial anxiety score, $F(19, 301{,}353) = 1{,}257.18$, $p < .001$, $R^2 = .073$. However, the model only accounted for a total of 11.3% of the variance in the depression scores, and 7.3% of the variance in anxiety scores. In both models Variance Inflation Factors for the independent variables ranged from 1 to 2.4, suggesting that multicollinearity was not beyond acceptable levels.

Coefficients for the Goldberg Depression score multiple regression model are presented in Table 4.21. After controlling for all other variables, previous history of depression was the strongest predictor of (higher) initial depression score ($B = 1.37$, $\beta = 0.268$, $t = 146.46$, $p < .001$). The next strongest predictors were (in order of magnitude): having a trade, certificate or apprenticeship (higher score); the number of depression treatments previously tried (higher score with more

treatments); having completed secondary education only (higher score); having only undertaken some or no secondary education (higher score); being referred by a health professional (higher score); and being from the United Kingdom (higher score).

The Goldberg Anxiety score model coefficients are displayed at Table 4.22. The strongest predictor of initial anxiety in the model, after controlling for all other independent variables, was being aged 20-29 years (B = 0.58, β = 0.138, t = 51.20, p < .001). The next strongest predictors were all associated with an increased anxiety score and included: number of depression treatments; history of depression; having a trade/certificate/apprenticeship; and being aged 30-39 years. Being located in Canada was the only variable which did not significantly predict baseline anxiety in the model.

Table 4.21. Multiple regression predicting initial Goldberg Depression score (n=340,313).

	B	SE_B	β
Gender			
Female	**-.054*****	.007	-.013
Age group			
** reference group: ≥ 50 years*			
≤ 19 years	**.089*****	.014	.015
20 – 29 years	**.120*****	.011	.030
30 – 39 years	**.074*****	.011	.017
40 – 49 years	**.107*****	.011	.022
Country			
** reference group: Australia*			
United Kingdom	**.254*****	.008	.067
United States	**-.098*****	.015	-.012
Canada	**-.102*****	.013	-.015
Location			
Rural/remote	**.031*****	.007	.007
Education level			
** reference group: Higher degree*			
Some secondary	**.468*****	.016	.069
Secondary only	**.465*****	.014	.075
Trade /certificate	**.431*****	.011	.099
Undergraduate	**.297*****	.015	.044
Diploma	**.343*****	.013	.060
Bachelor's degree	**.110*****	.011	.026
Health professional referral			
Referred	**.258*****	.007	.069
Depression history			
Previous depression	**1.37*****	.009	.268
Treatment history			
Number of treatments	**.062*****	.002	.075
Motivation level			
Motivated	**.061*****	.007	.014

$^{***}p < .001$

Table 4.22. Multiple regression predicting initial Goldberg Anxiety score (n=340,313).

	B	SE_B	β
Gender			
Female	**.218*****	.007	.051
Age group			
** reference group: ≥ 50 years*			
≤ 19 years	**.478*****	.015	.077
20 – 29 years	**.580*****	.011	.138
30 – 39 years	**.467*****	.011	.104
40 – 49 years	**.246*****	.012	.049
Country			
** reference group: Australia*			
United Kingdom	**.216*****	.008	.054
United States	**-.114*****	.016	-.014
Canada	.018	.014	.002
Location			
Rural/remote	**.053*****	.008	.012
Education level			
** reference group: Higher degree*			
Some secondary	**.516*****	.017	.073
Secondary only	**.483*****	.015	.074
Trade /certificate	**.497*****	.012	.109
Undergraduate	**.337*****	.016	.047
Diploma	**.381*****	.014	.063
Bachelor's degree	**.086*****	.012	.019
Health professional referral			
Referred	**.353*****	.007	.091
Depression history			
Previous depression	**.632*****	.010	.118
Treatment history			
Number of treatments	**.103*****	.002	.120
Motivation level			
Motivated	**.094*****	.008	.021

*** $p < .001$

4.4 Discussion

4.4.1 Summary of results

The results for each of the four aims of the study in this chapter are summarised below.

Aim 1: User characteristics

The spontaneous *MoodGYM* community users in this 5-year prospective study (2009-2014) were largely female (68.1%) and spanned all age groups, with the modal age being between 20-29 years (31.3%). Male users were significantly older than female users, and in the older age groups the relative rate of decline in users was less in men than women. Approximately half of all new users were from the United Kingdom (49.5%); a quarter (24.6%) resided in a rural or remote region. Users were highly educated, with 40.5% having completed a Bachelor's or higher degree. Of all users, 42.7% were referred to the program by a health professional, the majority (84.5%) reported a history of depression, and most had previously accessed evidence-based treatments for depression: 88.3% reported having accessed at least one treatment for depression and more than half (57.6%) at least three treatments. Half (50.6%) had used anti-depressant medication, and 42.6% did not report previously receiving a conventional depression treatment (anti-depressants, CBT, Interpersonal therapy or ECT). Nearly one quarter (24.7%) of users reported that they were committed to working through the whole program.

Aim 2: Inter-country comparisons

Most users (81.5%) were from the United Kingdom, Australia, Canada or the United States, and there were significant differences in the attributes of users from these different countries. Compared with other countries, a significantly greater percentage of users from the U.S. were male, and U.S. users were generally younger. Australian users were significantly more likely to be aged under 19 than U.K. users. The percentage of U.K. users from rural areas (30.3%) was significantly larger than for users from the other countries, and the lowest percentage of rural users were from Canada (16.3%). Compared with other countries, a significantly greater percentage of users from the U.K. reported a trade, apprenticeship or certificate qualification, and the largest percentage of users with university degrees (Bachelor's degree or higher) were from the U.S. (43.7%). Users from the U.K. were most likely to have accessed the program via referral from their GP or doctor (50.8%), whereas clicking on a web-link or online search was the most common pathway for U.S. users (62%). The percentage of users who reported to be committed to working through the program was significantly lower in Australia (23.3%) compared to the other countries.

Aim 3: Health professional referral and its predictors

Overall, 42.7% of users were referred to *MoodGYM* by a health professional. Controlling for other demographic characteristics, male users and users aged 40-49 years were more likely to be referred to the program by a health professional. Referral was also independently predicted by lower educational level, being from a rural or remote area, and being from the U.K. or Canada.

Aim 4: Initial symptom level and its predictors

Spontaneous community users had high initial Goldberg depression and anxiety scores registering a mean of more than six symptoms in each case, a level placing them at the upper end of the depressive symptoms range among members of the general community [143, p. 20]. There were significant differences in depression scores for users of different age groups, with the largest difference between users younger than 19 years and those older than 50 years, the latter showing lower initial scores. There was no significant difference in depression scores overall for male and female users. However, female users aged less than 19 had significantly higher depression scores than male users of the same age. Conversely, female users over 40 years had significantly lower scores than the males in their age cohort. Multivariate analysis showed that the strongest independent predictors of greater baseline depression symptoms were a previous history of depression, the number of evidence-based treatments that had been previously accessed, a lower education level (trade, secondary education or less), referral by a health professional and being from the U.K. The predictors of high baseline anxiety symptoms were similar; however, the strongest predictor was being aged 20-29 years, and being aged 30-39 years was also a strong predictor of initial anxiety.

4.4.2 User characteristics

The preponderance of female users identified in the current study (68.1%) is consistent with the findings previously reported in earlier studies of *MoodGYM* [130, 131, 132, 133] and also reflects the higher rate of depression in females compared to males in the general community [144]. Likewise, the higher percentages of female users in the younger age groups are in line with the earlier *MoodGYM* study of users aged 19 years or under [134], as well as evidence that relative rates of depression in women in the general community are highest in 14-25-year-old individuals [145]. The overall rate of female users was lower than that seen in the Spanish study of the *HDep* program, likely because *HDep* was designed for women [135]. The age profile of users was also similar to the previous studies: a large proportion of users in the current study were aged less than 30 years (41.8%) which is consistent with the mean user age of 35.5 years reported in the first *MoodGYM*

study [130] and to the finding in the 2008 *MoodGYM* study that 54% of users were aged less than 35 years [133]. The *HDep* study attracted slightly younger users, with 52.9% aged less than 31 years. The reason for this discrepancy is unclear, but at the time of the study Internet penetration was low in Mexico [146] (the primary source of *HDep* users), with the highest usage among young people [147].

Although there is little available evidence about the characteristics of individuals who access automated iCBT for depression, the finding (in both the current and previous studies) that a substantial percentage of *MoodGYM* users, as well as *HDep* users, were highly educated is consistent with previous research which has found that help-seeking behaviour for depression is positively associated with a higher level of education [148]. Higher education levels may reflect a higher level of literacy and concomitant increased exposure to information about depression with a resulting high level of depression literacy, including how to seek help.

The percentage of *MoodGYM* users who self-reported to be from a rural or remote region (24.6%) was consistent with the results from a previous study of 2006 *MoodGYM* users (22%) [133] and was higher than that among the general populations of the most represented countries. For example, 30.3% of U.K.-based *MoodGYM* users reported being from a rural area, but only approximately 17% of the U.K. population reside rurally [149]. Similarly, 21.1% of Australian users were from a rural or remote location compared to 10% in the Australian population. This suggests that individuals in rural areas, who have less access to face-to-face psychological therapies for depression [11], are more likely to access the program.

The current study found that most users have a history of depression (84.5%) and have tried evidence-based treatments for depression, including anti-depressant medication (50.6%). Hence, it could be that users typically have chronic or treatment refractory symptoms or are experiencing recurrent depression symptoms. This is consistent with evidence that recurrences of depression are common, even in individuals who have received treatment, including anti-depressant medication [150]. Recent research has found that face-to-face CBT results in reliable and clinically significant improvement in depression in approximately half of the individuals who receive it [151], and that half of those individuals will relapse within 1-2 years [152]. Interestingly, 20.1% of the current *MoodGYM* users indicated that they have previously used CBT for depression, with 71.0% finding it useful. It is possible that these users were experiencing a relapse of symptoms and were seeking a treatment that had assisted them in the past; others may have been seeking a treatment to supplement or substitute for other current treatments that were proving ineffective. In contrast, 42.6% of users had not accessed a 'conventional' depression treatment, which suggests that for

these users *MoodGYM* might have been a first-line treatment option rather than either an adjunct to conventional treatment or a second-line treatment.

There are several differences in the characteristics of *MoodGYM* users in the current study compared to earlier studies of the program. First, nearly half of users reported logging in from the United Kingdom (49.5%), a substantially larger percentage than in previous studies. Most users who registered in 2001 were from the United States (34.9%) or Australia (33.2%) [130], whereas those who registered in 2006 were mostly from Australia (41%) or the United Kingdom (31%) [134]. Hence the proportion of British users has increased over time, likely due to the integration of CCBT into the U.K. health system and the associated widespread promotion and awareness of iCBT programs such as *MoodGYM* in the U.K. In addition, the percentage of users in the current study who were referred to the program by a health professional (42.7%) is larger than for users who registered in 2006 (18%) [133]. This may be due, at least in part, to increased awareness of *MoodGYM* by health practitioners through media or scientific reports of positive effectiveness trials.

4.4.3 Cross-country comparisons

This is the first study to compare the characteristics of spontaneous, community users of automated iCBT for depression across different countries. We found both similarities and clear differences in the demographic profile of *MoodGYM* users from the four countries from which most registrants originated. In some cases, these patterns among *MoodGYM* users reflect the comparative similarities and differences in the overall demographic characteristics of the four countries. For example, in each of the countries the prevalence of depression is higher among women than men [145, 153, 154, 155]. It is therefore not surprising that female users exceeded male users in these countries. However, the percentage of users who were male was significantly higher in the U.S. than in other countries, a finding which does not reflect prevalence patterns. It may be that males in the U.S. were more likely than females to seek free iCBT self-help for mental health problems compared to male users in other countries. Alternatively, this result might be moderated by other factors, such as the higher education levels reported for U.S. users (see below).

Interpretation of differences in the education levels of users across countries is complicated by potential variability in the way that users reported their education level, with education systems using varying nomenclature. Notwithstanding this limitation, a larger percentage of users from the U.K. and Australia had a trade, apprenticeship or certificate, than users from Canada or the U.S. (which had the highest percentage of users currently attending university or having obtained a degree). How users self-defined being located in a ‘rural or remote’ area was also potentially

inconsistent across countries. For example, it is possible that Australian users living in regional centres selected 'rural or remote' rather than 'other city'. Despite this, it appears that smaller percentages of users from the U.S. and Canada were from rural areas. Thus, it is possible that in the U.K. and Australia *MoodGYM* was used by individuals from broader socio-economic backgrounds, which in turn may be linked to the pathway through which they found the program.

There were marked differences between countries in the pathways through which users accessed the service, which is likely due to differing health system contexts. For example, in the U.K., where CCBT is an established part of the health system, it is probable that many health professionals had knowledge of *MoodGYM* as an acceptable depression treatment and were recommending it to patients. Health professional referral is discussed further below. In contrast, in the U.S. where CCBT was not part of clinical guidelines or clinical policies, users were more likely to be highly educated and to discover the program by themselves through media stories or web searches.

It is also possible that the pathway through which a user found the program reflects the user's locus of control [156]. 'Self-referral' via a web link or Internet search may be associated with an internal health locus of control, characterised by a greater self-belief in one's capacity for personal control over one's health, whereas referral via a health professional may be more likely to signify an external locus of control, with a greater belief in the influence of external forces on one's health. Although there is no existing literature which points to a collective difference in locus of control for citizens of these different countries, research does suggest that being located in different communities within the same country can foster differences in locus of control [157]. Similarly, individuals residing in rural areas often score higher on measures of self-reliance and stoicism [158], which could suggest greater internal locus of control among this group. Future research into the health locus of control of community users of iCBT would improve our understanding of what types of users engage with the service, and how to promote publicly available interventions to different groups of potential users.

4.4.4 Referral by health professionals

To our knowledge, there have been no studies to date of the user characteristics which predict referral to iCBT by health professionals. A range of studies have investigated factors which influence clinicians' referral of individuals to specialist mental health care for depression [e.g., 159]; however, it is likely that factors which predict referral to automated iCBT self-help would be quite different and, for example, would exclude individuals with high levels of depression.

In the current study of *MoodGYM* user characteristics, after controlling for other demographic variables, users from the U.K. were more likely to have been referred by a health professional, a

finding that, as discussed above, might be explained in large part by the health system context in the U.K. Being from Canada also predicted health professional referral. It is unclear why. However, we note that the 2008 *Family Physician Guide* for depression issued by the British Columbia Ministry of Health included *MoodGYM* in its resources section [160].

Significantly, referral to *MoodGYM* by a health professional was predicted by lower educational level, after controlling for other variables including country. This is a somewhat surprising finding given that a priori it might be expected that the verbal and cognitive demands of *MoodGYM* could render it more suitable for those with a higher educational level. (For example, lower educational level was associated with lower intervention adherence in an individual patient data meta-analysis of studies of unguided iCBT for depression [37]). However, it is possible that users with lower educational levels were also of lower socio-economic status in which case it might be that *MoodGYM* was an option provided to the patient by their doctor because it was free (or chosen by the user from amongst options presented by the health practitioner because it was free).

Being male also predicted referral by a health professional, which is also counter-intuitive given that men seek help for depression from professionals at half the rate of women [161]. It is not known if practitioners were more likely to recommend *MoodGYM* to men than to women, or if men were more likely than women to choose *MoodGYM* from the options presented to them by their doctor. It is possible that the anonymous nature of the program, which avoids the stigma of accessing face-to-face help, was more often seen as a preferred treatment option for men by health professionals, by the men themselves, or by both given that the stigma associated with depression is higher among men [162]. Alternatively, males might have been less likely to actively seek avenues of self-help for depression unless specifically directed to do so by a professional. In order to better understand this finding, further research is required on the attitudes of health professionals to suggesting iCBT to patients, and on gender differences in responding to iCBT self-help recommendations from health professionals.

More expected was the finding that users located in rural or remote regions were more likely to have been referred by a health professional than users from urban areas. There is qualitative evidence from Australia that GPs in rural areas identify cost, physical distance, long waiting lists, stigma and lack of knowledge about resources as barriers to accessing mental health services [163]. Hence, for rural users, doctors may have utilised e-mental health as a free, accessible solution to these barriers and 'prescribed' *MoodGYM* in the absence of the availability of other services [11].

4.4.5 Depression and anxiety symptoms of users

Although the mean level of symptoms reported in all *MoodGYM* community user studies is higher than those in the general population [143, p. 20], the mean depression and anxiety symptoms of users of Version 3 of *MoodGYM* in the current study were higher than previously reported symptom levels for community users of either Version 1 or 2 of the program [132]. Likewise, the *HDep* study undertaken between 2009 and 2013, a similar time period to that of the current study, reported relatively high levels of depression in users, with the majority meeting the clinical cut-off for depression on the CES-D scale [135]. This increase in the severity of depression among users over time may reflect a growing awareness of the effectiveness of online interventions for clinical depression, especially among health practitioners (see below). Whatever the explanation, the high level of symptoms among *MoodGYM* users suggests that the service is reaching individuals with a high level of need.

Although previous studies have reported on the level of depressive symptoms among users, no previous study of spontaneous community users has examined the demographic predictors of initial depression or anxiety scores.

A self-reported history of depression was the strongest predictor of both initial depression and anxiety symptoms, after controlling for other variables including age and gender. It is possible that a history of depression was associated with chronic or treatment-resistant symptoms (and hence higher scores), and that users without prior depression were more likely to use the program for prevention or early intervention (and hence exhibited lower depression scores).

Users who were referred by a health practitioner also had higher symptoms. This may reflect both that the symptoms were of sufficient severity and impact to trigger face-to-face help-seeking and that in the clinician's expert opinion – based on either formal or informal assessment – the user required a depression intervention. A previous Swedish study of iCBT for depression investigated differences in baseline depression scores for participants as a function of their recruitment pathway [164], albeit in the context of a formal randomised controlled research trial rather than an open-access intervention. The authors found no difference in the scores for participants who were referred to the trial by a 'clinical pathway' compared to those recruited by online search. However, in the latter trial, clinical pathway included not only referral by a health professional but also referral via web-links from mental health, medical or well-being websites, from other trials of iCBT, and from websites which advertised online trials. The baseline depression severity of the subset of participants who were referred by health professionals was not analysed separately. Thus, it is not possible to

exclude the possibility that the severity depression was higher for participants who were referred by health professional compared to other users. Interestingly, in the current predictor analysis, even after controlling for referral by a health practitioner, being from the U.K. predicted higher symptom scores. This suggests that general awareness of iCBT was higher in the U.K., and hence people who needed treatment for depression and anxiety were more aware of *MoodGYM.*

Higher symptoms were also strongly predicted by lower education levels; users with higher education levels had *lower* symptoms. This finding raises the possibility that users who were more highly educated were most interested in learning skills for prevention or early intervention, whereas *MoodGYM* was more likely to be a treatment option for less educated users.

Interestingly there was no significant difference in the initial depression scores of female compared to male users. However, young female *MoodGYM* users and males aged 40-49 had the highest mean depression scores. This could point to an unmet need for services in individuals with high levels of symptoms in these sub-groups, or potentially a greater preference among these groups for anonymous online self-help over other forms of depression treatment. Likewise, the difference in symptom levels for female (higher) vs male (lower) users aged 19 years or younger could reflect the much higher rate of depression in young women compared to men [144] or could suggest that automated iCBT has high acceptability for young females with high levels of depressive symptoms.

After controlling for all other variables, being aged 20-29 years and 30-39 years were also strong predictors of baseline anxiety levels (higher scores). This is consistent with evidence that the prevalence rates for Generalised Anxiety Disorder (GAD), Social Anxiety Disorder (SAD) and specific phobia are highest in individuals aged 18 to 35 years [165]. The findings are also consistent with the high acceptance of the use of technology for health management in individuals of this age [e.g., 166].

4.4.6 Limitations

There are several limitations associated with the current study.

First, this study was an exploratory analysis with no *a priori* hypotheses. In addition, many variables were intercorrelated. For example, there was a positive correlation between initial depression score and history of depression, $r = .29$, $p < .001$, $n = 372{,}422$; and a positive correlation between history of depression and number of treatments accessed, $r = 0.34$, $p < .001$, $n = 417{,}334$. Hence, it is difficult to ascertain with confidence which might be the most important factors. Further, the large size of the dataset means that even small differences in magnitude can be highly statistically significant; thus

inferences must be approached with caution [167]. However, conversely, this does mean that differences and relationships which are not significant can be confidently categorised as such.

Another limitation is that the data collected may have omitted key variables which could be important in understanding users of iCBT for depression. For example, the current data set did not include some potentially relevant demographic measures such as socio-economic status, level of employment and relationship status. In addition, social and behavioural factors such as level of social support, concurrent help-seeking and locus of control were not assessed.

In addition, in the absence of a suitable measure, we cannot exclude the possibility that some users of *MoodGYM* undertook the program with external guidance. A large proportion of users indicated that they were referred to the program by a health professional; however, there is no data indicating whether users were supported by their practitioner in the use of the program. Any user who received support to register and engage with the program would have been accessing a guided iCBT intervention rather than automated self-help.

Finally, the data analysed in the current study is self-reported by anonymous, global users, with no means to confirm its accuracy.

4.5 Conclusion

This study demonstrates that *MoodGYM* is a viable public health intervention, given the large number of public users who accessed the program over approximately five years. These users were predominantly from the United Kingdom, Australia, Canada and the United States, with the most users coming from the U.K. where iCBT has been included as part of clinical guidelines for depression and promoted through national health services. This suggests that the success of automated iCBT as a public health intervention is facilitated by health system factors that promote clinician and community awareness of evidence-based technology interventions for mental health. Users had high levels of symptoms, and a large proportion was referred to the program by a health professional, suggesting that the program met a broader need for free, accessible depression treatment.

Further research is required to refine and extend our understanding of the characteristics of users who access the program via different pathways, including potential relationships with the users' locus of control. Additionally, more research is required to explore the drivers of referral to automated iCBT by health professionals. The results of this study showed that users were more likely to be referred if they were male, from a rural or remote region or had lower educational attainment, but it is not clear if these associations were due to user preferences or the attitudes of practitioners

and their decisions about for whom unguided iCBT is appropriate. Finally, although the program was used by both male and female users from a wide age range, further research is required to better understand attitudes and expectations of iCBT in sub-groups of the community, in order to more effectively promote the availability and/or appropriateness of free, automated iCBT to those for whom it might be helpful.

5. MoodGYM Spontaneous Community Users: Adherence and depression outcomes

5.1 Introduction

The previous chapter demonstrated the high level of reach that can be achieved by an online mental health intervention. However, the public health impact of such programs is also dependent on the effectiveness of intervention (e.g., whether participant depression symptoms reduce over time), which in turn may depend on how much users engage with a program (i.e., adherence to the intervention). Real-world observational studies of adherence and effectiveness are required since, as noted in the previous chapter, evidence from automated iCBT for depression from RCTs may not translate into equivalent outcomes for individuals accessing the program outside of research contexts [125].

With respect to user engagement with iCBT programs, research trials offer participants structure for their use of an intervention (for example they are 'followed up' for post-test measures) and elicit an initial commitment to ongoing involvement in the study, both of which are likely to enhance a user's adherence to the program [168]. In contrast, spontaneous users of publicly available automated iCBT programs may be less motivated to engage with an iCBT program. If an automated program is freely available, users who register may be 'just having a look' at what is offered, having made no commitment to participate and bearing no financial cost if they do not continue. This is supported by a systematic review published in 2012 of studies involving Internet interventions for health conditions which found that users of Internet interventions who were part of observational studies had considerably lower adherence levels than RCT participants [169].

The depression outcomes for spontaneous users may also differ from those reported for research participants. However, this is difficult to assess due to the nature of data collected in real-world contexts, in which spontaneous users are typically not followed up to complete post-test measures after the elapse of a predefined time period. Rather, investigations of changes in symptom levels for spontaneous users rely on survey responses provided as part of the intervention. As such, repeated assessments are collected only for those users who sufficiently engage with the program to undertake subsequent surveys; further, across individuals, this repeated assessment data will represent responses at varying time intervals from intervention commencement. In addition, depression outcome measures comprise scales embedded in interventions which may be shorter or less clinically focused than those routinely used in RCTs. Notwithstanding these limitations, analysis of change in depression survey responses for spontaneous users may provide insight into the

changes associated with iCBT use for those individuals who engage with them, and into what type of users may benefit most from publicly available iCBT interventions.

As reported in the previous chapter, a recent systematic review of observational or implementation studies of unguided Internet interventions for depression or anxiety identified six real-world studies of iCBT for depression [127]. Five studies reported on spontaneous users of the *MoodGYM* program (Versions 1 and 2) [130, 131, 132, 133, 134], and the adherence and symptom outcome results reported in these studies will be described in the next section. In addition, one study reported on 17,318 community users of the *HDep* Spanish-language program [135]. Of the users in the latter study, 16,564 (95.6%) completed the first depression survey (CES-D), and 12,366 (71.4%) accessed the first of seven *HDep* modules. Module completion rates declined rapidly over the seven consecutive modules with 42.1%, 20.5% and 3.6% of users accessing Modules 2, 3, and 7, respectively. Predictors for accessing at least one, two or three modules included: 1) female gender; 2) age over 30 years; 3) self-reported disability due to depression; 4) self-reported previous suicide attempt; 5) being a homemaker; and 6) being employed. Note that the program was designed for women, and hence it is unsurprising that female gender was a predictor for adherence to the intervention in each of these models. Only 592 users (3.4%) completed the intermediate CES-D assessment after Module 3, and 191 users (1.1%) completed the final CES-D assessment. Neither changes in CES-D scores for these users nor their predictors were reported by the authors.

5.1.1 Previous reporting of adherence and depression outcomes for spontaneous, community users of MoodGYM

Adherence and outcome data from previous studies of spontaneous users of *MoodGYM* are summarised in Table 5.1 and are described below. All of the five studies analysed adherence rates and three evaluated depression outcomes.

Table 5.1. Adherence and depression outcomes reported in studies of spontaneous MoodGYM community users.

Study	Registration period, *N*	Adherence	Outcomes	
			Goldberg Depression changes	**Goldberg Anxiety changes**
Christensen H, et al., 2002 [130]	April 2001 – Sept 2001 *N*=2,909	51.7% ≥ 1 DA; 16.0% ≥ 2 DAs.	Improvement predicted by more completed modules, more than 1 week between assessments	Improvement predicted by more completed modules, more than 1 week between assessments
Christensen H, et al., 2004 [131]	April 2001 – Sept 2003 *N*=19,607	61.9% ≥ 1 DA; 15.6% ≥ 2 DAs Females more likely than males to complete 2 DAs.	Improvement predicted by more completed modules (up to 4) but not gender (repeated measures ANOVA)	Improvement predicted by more completed modules (up to 4).
Christensen H, et al., 2006 [132]	April 2001 – Oct 2004 *N*=58,398 V1: *n*=19,607 V2: *n*=38,791	V1: as above. V2: 51% ≥ 1 DA; 18% ≥ 2 DAs	V1: Lower final score predicted by: female; more completed modules; lower initial depression score.	V1: Lower final score predicted by: female; more completed modules; lower initial anxiety score.
Batterham PJ, et al., 2008 [133]	Jan 2006 – April 2007 *N*=82,159	57.7% ≥ 1 DA; 20.5% ≥ 2 DAs More modules predicted by: younger age, higher education, European location, referral by health professional, female, higher initial depression score.	NR	NR
Neil AL, et al., 2009 [134]	Jan 2006 – Nov 2007 (≤ 19 years) *N* = 7,207	89% ≤ 1 module More modules predicted by: female, higher initial depression score.	NR	NR

DA= Depression Assessment; NR= Not reported; V1= Version 1; V2= Version 2.

The first study of spontaneous *MoodGYM* users, published in 2002, reported on 2,909 users who registered in the first 6 months of the public delivery of the program from April 2001 [131]. Of these users, 1,503 (51.7%) completed at least one Goldberg Depression scale and 465 (16.0%) completed at least two depression assessments. For users who completed at least two modules, changes in depression and anxiety scores (last depression or anxiety score recorded minus first depression or anxiety score recorded) were significantly predicted by the number of modules completed, with more completed modules associated with greater improvement. Larger improvement was also associated with more time between assessments (over a week between first and last measures, compared to completion within a week or assessments completed on the same day).

A follow-up study investigated 19,607 users (including the users from the previous study) who registered over approximately 30 months from April 2001 [131]. The majority of users completed at least one depression assessment (*n* = 12,141, 61.9%); however only 15.6% (*n* = 3,055) completed at least two depression assessments, and female users were significantly more likely than male users to complete at least two assessments. This study also compared these community users to 182 participants from an RCT of *MoodGYM* and found that the RCT participants had higher adherence to the intervention with 86.3% completing one depression measure and 66.5% completing two or more assessments. Based on analysis of those users for whom there were data for at least two time-points (*n* = 3,176 for depression, *n* = 1,688 for anxiety), greater improvement in both depression and anxiety scores were predicted by an increased number of modules completed (from two to four modules), but no significant improvement was predicted by completion of five compared to four modules. After controlling for the number of completed modules, there was no difference in change scores for community users compared to trial users.

The next study included both the 19,607 community users from the previous study who registered on Version 1 of the program and an additional 38,791 users who registered on Version 2 between September 2003 and October 2004 [132]. There was a significant difference in the number of depression tests completed for Version 1 compared to Version 2 users. Approximately half (51%) of Version 2 users completed at least one depression assessment (compared to 61.9% of Version 1 users[5]) and 18.0% completed two or more assessments (compared to 15.6% of Version 1 users). Final depression and anxiety scores were analysed for the 3,176 and 1,688 Version 1 users with repeated depression and anxiety data respectively, and lower final scores were predicted by being female,

[5] The paper for this later study reported that 46% of the Version 1 users completed at least one depression assessment [132]; however, the original study reported that 61.9% of the same sample of Version 1 users completed at least one depression assessment [131] and it is this latter figure that we report here.

having a lower initial depression/anxiety score and by more modules completed. No final score data were reported for Version 2 users.

An additional two studies focused on predictors of adherence for community users of *MoodGYM.* The first of these examined 82,159 users who registered on Version 2 of the program from January 2006 to April 2007 [133]. A nominal logistic regression model was constructed to analyse completion of either none, one, or at least two modules, based on registration and assessment data from the 59,453 users who completed the initial depression, anxiety and dysfunctional thoughts surveys in the program. After controlling for other independent variables, users with higher initial depression and dysfunctional thinking scores were significantly more likely to have completed more modules, and higher initial anxiety was associated with higher odds of completing one module, but not two or more modules. Completion of more modules was also significantly associated with younger age groups and being referred to the site by a health professional. Users with higher education levels were also more likely to complete two or more modules. There was a small significant gender effect, with female users more likely to complete two or more modules. Being located in a rural or remote region did not predict module completion. Users from European countries were more likely to complete modules, whereas users from North America were less likely to complete modules. Of the users who completed at least one module, completion of two or more modules was more likely for those whose depression scores either improved or remained stable after the first module, compared to those whose scores worsened.

A second study examined community users who registered with *MoodGYM* between January 2006 and November 2007 and who self-reported to be aged 19 years or under (N = 7,207) [134]. This study also reported on 1,000 students who participated in a supervised school classroom based RCT of *MoodGYM*. Most of the community users (89%) completed no or one module, whereas most users in the school sample (85%) completed three or more modules. A multiple linear regression analysis of data for the combined sample of community users and RCT participants demonstrated that adherence was predicted by being an RCT participant rather than a community user, and by being female. For those in the community sample, higher initial depression symptoms (but not anxiety symptoms or level of dysfunctional thinking) predicted adherence.

Limitations of existing studies

As noted in Chapter 4, the existing studies of community users of *MoodGYM* are now relatively old, having been conducted within the first decade of iCBT delivery. In particular, analysis of predictors of outcome for community users has been conducted solely on a sample of approximately 19,000 users of the first version of the program who registered between 2001 and 2003. Given the previously

noted expansion of research on iCBT, increased public access to the Internet and the adoption of e-mental health approaches by governments and health systems, studies of more recent users are required to update the evidence-base to ensure that it is relevant to the current public health context.

Another limitation of the extant literature is that the studies of the predictors of adherence and outcomes have been largely limited to the analysis of demographic and clinical variables. However, other variables may play a role in adherence behaviour and symptoms outcomes. One potentially relevant variable in this context is the motivation of the user at the time of registration to engage with the program. The (lack of) motivation of users is often cited as a potential barrier to adherence to self-guided interventions [e.g., 170], and several studies have found that motivation (and specific motivators) are important to user persistence with iCBT in both self-guided and guided modalities [119, 171, 172]. A user's attitude or expectation towards CBT could also potentially influence their use of the program. For example, a study of predictors of depression outcome response for a guided iCBT intervention found that perceived credibility of iCBT and patients' expectations were strong predictors of improved outcomes [173]. However, there are no studies which focus on the effect of motivation or attitudes for spontaneous community users of iCBT programs. An understanding of factors that predict outcome may assist in identifying those users who are most likely to benefit from an iCBT program and in the case of potentially modifiable factors (such as CBT literacy and initial motivation) to tailor pre-interventions to individuals with the aim of improving the iCBT outcomes.

5.1.2 Study aims

The current study addresses these limitations and extends previous studies by examining adherence and outcomes among spontaneous community users who registered on Version 3 of the *MoodGYM* program, over approximately 5 years from March 2009. It has two aims:

Aim 1: *To investigate adherence among spontaneous MoodGYM users as measured by user progress through the program, and whether progress can be predicted by demographic and clinical characteristics, motivation level or previous experience of CBT.*

Aim 2: *To investigate changes in self-reported depression symptoms of spontaneous MoodGYM users, and whether symptom changes can be predicted by user characteristics, including motivation level, previous experience of CBT, or adherence to the program.*

5.2 Method

The data examined in this study were from the same user dataset reported in Chapter 4. The study received approval from the Australian National University Human Research Ethics Committee (Protocol Number 2013/592).

5.2.1 Participants

Eligible participants were spontaneous *MoodGYM* users who registered with the English language open-access version of the *MoodGYM* program between 6 March 2009 and 6 January 2014 (*N* = 417,354). As per the inclusion and exclusion criteria for the study described in Chapter 4, these were 'personal' community users who indicated that they were seeking help for themselves.

5.2.2 Measures

Registration data

Registration data were collected via a self-report at the time of registration. This data, which is detailed in Chapter 4, comprised gender, age group, country of registration, rural/urban location, education level, pathway to *MoodGYM*, previous history of depression, treatment history, previous experience of CBT and motivation level.

After registration, users were required to complete the initial Goldberg Depression and Anxiety scales and the Warpy Thoughts Quiz before proceeding to the first module.

Depression and anxiety symptom levels

As described in the previous chapter, responses to the depression and anxiety assessments produced a score from 0 to 9, with higher scores representing more symptoms [136]. The depression and anxiety scales were repeated at the start of each of the subsequent modules, and users were required to complete these measures in order to progress through the program. Although participants could review and repeat completed surveys, only their initial responses to outcome measures were recorded at each module.

The post-test depression score for a user was defined as the repeated measure recorded at the most advanced point in the program. The change in depression symptom levels was calculated as the initial score minus the post-test score so that a positive change reflected an improvement in depression symptoms.

Level of dysfunctional thinking

The 42-item Warpy Thoughts Quiz is a scale that was developed for the *MoodGYM* program to measure and provide feedback about users' dysfunctional thinking patterns [137] and was administered as part of the initial assessments before Module 1, and then repeated half-way through Module 2 and at the end of the intervention. In the program, the scale is divided into seven areas of dysfunctional thinking: the need for approval; the need to be loved; the need to succeed; perfectionism; influence on others; external requirements for happiness; and expectations of rights. Each section has six dysfunctional thinking statements, with possible responses on a 5-point scale from 'strongly agree' to 'strongly disagree'. However, psychometric evaluation has demonstrated the measure has three first-order factors which are based on 20 of the 42 items – Relationships (ten items), Achievements (five items) and Entitlements (five items), as well as a higher order factor, 'Warpy Thoughts' which corresponds to all 20 items [137]. The current study employed the latter, and a Warpy Thoughts score was calculated based on the responses to the 20 statements identified in the factor analysis. Responses were reverse coded, from 4 'strongly agree' to 0 'strongly disagree' and summed to give a Warpy Thoughts score out of 80, with higher scores representing higher levels of dysfunctional thinking.

Adherence and related measures

Data indicating user progress and level of interaction with the program were extracted from the *MoodGYM* database as follows:

Number of modules: The number of completed modules (0 to 5) for each user was extracted, with a module deemed completed if the user progressed to the final summary page for that module.

Number of sessions: The number of sessions or times that a user logged in to access the program was extracted. More than ten logins were coded into a discrete category.

Estimated time on program: The amount of time that a user spent on program webpages was calculated from the program weblogs. This value represented an estimate of time spent on the intervention, since a user may have technically stayed logged in and on a certain page but have moved their attention elsewhere. This data was coded into four discrete categories: less than 30 minutes; between 30 and 60 minutes; between 60 and 120 minutes; and 120 minutes.

Time elapsed between pre- and post-tests: The elapsed time between user completion of the initial depression assessment and their post-test assessment was calculated and categorised into: within 24 hours; 24 hours to 1 week; more than 1 week.

5.2.3 Analyses

All statistical analyses were conducted using SPSS Version 25.0 [140].

Aim 1: Adherence and its predictors

Program starters and non-starters

The percentage of *MoodGYM* registrants who completed the first assessment phase (initial symptom and dysfunctional thinking measures) and the percentage who progressed to commence the first module ('starters') were computed.

Separate Chi-square tests were conducted to investigate the association between starter status (starters vs non-starters) and each of the registration variables for users from one of the top four countries by usage (United Kingdom, Australia, Canada, United States). For variables with more than two categories, pairwise post hoc tests were performed using a Bonferroni correction to account for the type 1 error associated with multiple comparisons. A logistic regression was then undertaken to explore whether 'non-starter' status was independently predicted by each of the registration variables. The dependent variable was 'Started Module 1' (Yes, No), and the independent variables included in the model were:

- Gender: female [0, 1];
- Age group: dummy coded with ≥ 50 years as the references group, ≤ 19 years [0, 1], 20-29 years [0, 1], 30-39 years [0, 1], 40-49 years [0, 1];
- Country: dummy coded with Australia as the reference group, U.K. [0, 1], Canada [0, 1], U.S. [0, 1];
- Location: rural [0, 1];
- Education level: dummy coded with Higher degree as the reference group, Some/no secondary [0, 1], Secondary only [0, 1], Trade/certificate [0, 1], Undergraduate [0, 1], Diploma [0, 1], Bachelor's degree [0, 1].
- Referral by a health professional: referred [0, 1];
- Previous history of depression: depression history [0, 1];
- Motivated to complete the program: motivated [0, 1];
- Previous experience of CBT: dummy coded with 'Not sure what CBT is' as the reference group, 'Tried and found it useful' [0, 1], 'Tried but not useful' [0, 1], 'Not tried but want to' [0, 1], 'Not tried and don't want to' [0, 1].

- Number of treatments previously tried: treatments [scale from 0 to 12].

All independent variables were entered simultaneously into the regression model. To ensure that the model was not confounded by multicollinearity, the relationships between the independent variables were checked by regressing each independent variable against the other independent variables and inspecting the Variance Inflation Factors (VIFs) for the variables in each of the linear regressions [174].

Number of modules completed

The number of modules completed by users from the top four countries who commenced Module 1 was examined, and the registration characteristics and initial symptom levels compared for those users who completed modules, one module or two or more modules. Collapsing completion of two, three, four and five modules into one category was undertaken since relatively few users (5.4%) completed three or more modules. In addition, since this trichotomous module completion variable was used in a previous study of *MoodGYM* community users [133], its use facilitated a direct comparison between the current and previous results. Further, most of the CBT concepts within *MoodGYM* are introduced from Module 2.

Univariate comparisons were undertaken for these groups, with Chi-square statistics calculated for categorical variables, and pairwise post hoc comparisons (corrected for multiple comparisons) undertaken for variables with more than three categories. One-way ANOVAs were performed for continuous measures, and Levene's tests were undertaken to test the assumption of homogeneity of variance. Given the large sample sizes, this assumption was violated in all cases, and Welch's *F* was calculated for each analysis. Games-Howell post hoc comparison tests were also undertaken to compare mean scores between each category of module completion.

A cumulative ordinal logistic regression was conducted to explore which demographic or other registration characteristics, initial symptom scores or dysfunctional thinking levels independently predicted completion of zero, one or two or more modules. This type of regression assumes that each independent variable has an identical effect at each cumulative split of the dependent variable. However, a full likelihood ratio test showed that this assumption of proportional odds was violated. Therefore, instead, a multinomial logistic regression was undertaken. The dependent variable was 'Number of modules completed' (0, 1, 2+) with no competed modules (0) as the reference category. The independent variables were those listed above for the previous binary logistic regression, as well as the user's initial Goldberg depression and anxiety scores, and initial Warpy Thoughts score. In order to check for multicollinearity, a series of linear regressions were conducted on each of the

independent variables in the model, against the other independent variables, to inspect for VIFs [174].

Number of logins and estimated time on the program

The number of logins and estimated time on the program were examined for users who commenced Module 1, including the subset of users from each of the top four countries.

Aim 2: Depression symptom changes

A paired samples *t*-test was conducted to evaluate the differences in initial depression assessment scores (d_i) and post-test scores (d_p), for those users who completed at least two depression assessments. Users who did not complete at least two depression assessments were not included in the analysis. The Cohen's *d* effect size was calculated based on the difference in means and the pooled standard deviation at pre- and post-test [$(d_p - d_i) / SD_{pooled}$], with a positive effect size signifying a reduction in symptoms over time.

To investigate whether a 'dose-response' relationship was present, univariate analyses were conducted to compare the effects of the adherence variables on change in depression score, after controlling for initial depression score. Correlation between adherence variables was also examined. In addition, univariate analyses were conducted to explore the relationship between the users' registration characteristics and change in depression score.

Hierarchical regression analysis was then applied to examine the extent to which depression scores could be explained by the independent variables. The variables were entered in blocks with baseline depression, anxiety and Warpy Thoughts scores entered in the preliminary control block, followed by demographic and other registration variables, and adherence variables as the final block.

5.3 Results

5.3.1 Aim 1: Adherence and its predictors

Program starters and non-starters

The number of registrants who completed the initial assessment surveys and who continued to Module 1 is displayed in Figure 5.1. Of the 417,354 registrants, 372,450 (89.2%) completed the initial depression assessment survey, but only 81.4% (*n* = 339,756) progressed through the initial surveys to the start of Module 1.

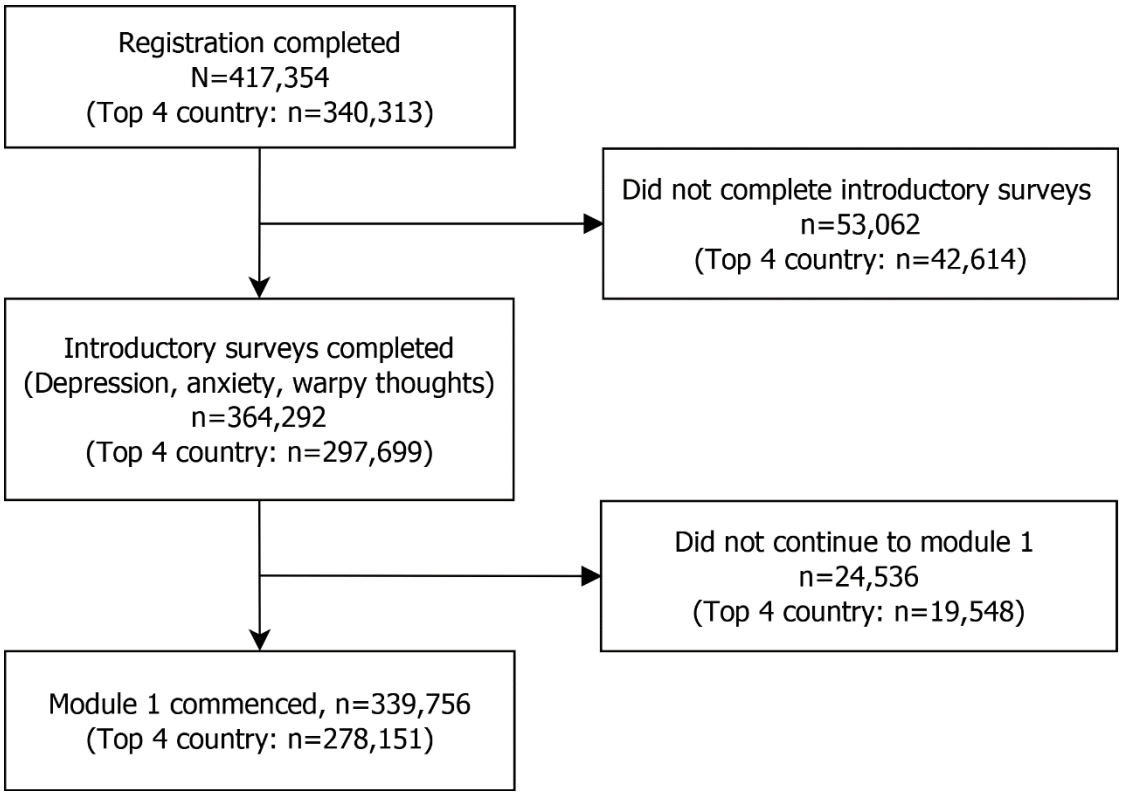

Figure 5.1. Number of users who completed introductory surveys and commenced Module 1 (N=417,354 registrants).

Table 5.2 shows the numbers of users who registered from one of the top four countries by usage (*n* = 340,313), who started and did not start Module 1 ('starters' and 'non-starters'). Univariate analyses showed statistically significant differences in the proportion of starters and non-starters for each of the registration variables. In particular, users with the following characteristics at registration were more likely to proceed to Module 1: female gender, referred by a health professional, from a rural area, a history of depression, and self-reported motivation to complete the program. There were also significant associations between starter status and age group (a lower proportion of registrants aged 50 years or over started Module 1), country of registration (a higher proportion of users from the United Kingdom started Module 1), education level (a higher proportion of undergraduate users were starters), previous experience of CBT for depression (a higher proportion of users started Module 1 who self-reported that they had not previously tried CBT for depression but wanted to try it), and number of treatments previously tried (a lower proportions of users who had tried a large number of treatments were starters).

Table 5.2. Whether or not users from the top four countries started Module 1, by registration characteristics (n=340,313).

	Started Module 1 (column %)	Did not start Module 1 (column %)	Chi-square	Cramer's V
Gender			**103.11*****	0.017
Female	193,979 (30.3%)	42,060 (67.7%)		
Male	84,172 (32.3%)	20,102 (69.7%)		
Age group			**1025.26*****	0.055
≤ 19 years[a]	30,428 (10.9%)	6,944 (11.2%)		
20 – 29 years[b]	87,182 (31.3%)	16,839 (27.1%)		
30 – 39 years[c]	69,635 (25.0%)	14,770 (23.8%)		
40 – 49 years[a]	52,246 (18.8%)	12,239 (19.7%)		
≥ 50 years[d]	38,660 (13.9%)	11,370 (18.3%)		
Country			**510.75****	0.039
United Kingdom[a]	171,384 (61.6%)	35,314 (56.8%)		
Canada[b]	21,420 (7.7%)	5,689 (9.2%)		
United States[b,c]	15,854 (5.7%)	3,976 (6.4%)		
Australia[c]	69,493 (25.0%)	17,183 (27.6%)		
Location			**43.81*****	0.011
Rural / remote	73,375 (26.4%)	15,596 (25.1%)		
Urban	204,776 (73.6%)	46,556 (74.9%)		
Education level			**615.07*****	0.043
Some secondary[a]	22,222 (8.0%)	6,341 (10.2%)		
Secondary only[b]	27,517 (9.9%)	5,933 (9.5%)		
Trade /certificate[c]	65,278 (23.5%)	15,556 (25.0%)		
Undergraduate[d]	22,732 (8.2%)	4,130 (6.6%)		
Diploma[c]	33,137 (11.9%)	7,832 (12.6%)		
Bachelor's degree[b]	70,137 (25.2%)	14,512 (23.3%)		
Higher degree[b]	37,128 (13.3%)	7,858 (12.6%)		
Health professional referral			**86.58*****	0.016
Referred	135,233 (48.6%)	28,940 (46.6%)		
Not referred	142,918 (51.4%)	33,222 (53.4%)		
History of depression			**82.74*****	0.016
History	235,121 (84.5%)	51,632 (83.1%)		
No history	43,030 (15.5%)	10,530 (16.9%)		

Table 5.2. Whether or not users from the top four countries started Module 1, by registration characteristics (n=340,313) (continued).

	Started Module 1 (column %)	Did not start Module 1 (column %)	Chi-square	Cramer's V
Motivation level			**316.15*** **	0.030
Motivated	69,547 (25.0%)	13,437 (21.6%)		
Not motivated	208,604 (75.0%)	48,725 (78.4%)		
Experience with CBT			**1275.06****	0.062
Tried, useful[a]	38,302 (13.9%)	9,990 (16.3%)		
Tried, not useful[b]	16,836 (6.1%)	4,100 (6.7%)		
Not tried, want to[c]	148,797 (54.0%)	28,381 (46.3%)		
Not tried, don't want to[d]	8,971 (3.3%)	2,733 (4.5%)		
Not sure what CBT is[a]	62,426 (22.7%)	16,072 (26.2%)		
Previous treatments			**393.78****	0.034
0[a,b]	31,481 (11.3%)	7,415 (11.9%)		
1[b.c]	39,778 (14.3%)	8,870 (14.3%)		
2[d]	47,205 (17.0%)	9,868 (15.9%)		
3[d]	45,484 (16.4%)	9,348 (15.0%)		
4[c,d]	38,228 (13.7%)	8,155 (13.1%)		
5[b,c]	29,929 (10.8%)	6,712 (10.8%)		
6[a,e]	21,077 (7.6%)	5,173 (8.3%)		
7[a,b]	13,235 (4.8%)	3,245 (5.2%)		
8[e,f]	6,991 (2.5%)	1,860 (3.0%)		
9[f,g]	3,241 (1.2%)	975 (1.6%)		
10[g]	1,172 (0.4%)	409 (0.7%)		
11[f,g]	292 (0.1%)	106 (0.2%)		
12[g]	38 (0.01%)	26 (0.04%)		
Total	**278,151 (81.7%)**	**62,162 (18.3%)**		

***$p < .001$.

For variables with more than two categories, each superscript letter denotes a subset of variable categories whose column proportions do not significantly differ from each other (after correction for multiple comparisons).

A logistic regression analysis was conducted to investigate which registration variables independently predicted commencement of Module 1. Independent variables were added to the regression analysis simultaneously. The model was statistically significant, χ^2 (23) = 3,492.96, $p < .001$); however, Nagelkerke pseudo R^2 showed that the model accounted for only 1.7% of the variance in whether

users started the first module (see Table 5.3). In order to check for multicollinearity, a series of regression analyses were undertaken for each of the independent variables against the other independent variables. All VIFs were within acceptable limits (the highest VIF was 1.23), suggesting that multicollinearity was not an issue in the model. The Hosmer and Lemeshow test statistic was significant (p = 0.03), indicating that there was a significant difference between the predicted and observed values, suggesting a poor model fit. However, this measure of model fit is unlikely to be reliable given the large size of the dataset [142].

Table 5.3. Logistic regression model test statistics.

	Chi-square (*df*, *p*)	**-2 Log likelihood**	**Cox & Snell R^2**	**Nagelkerke R^2**
Model (gender, age, country, location, education, referral, previous depression, motivated, experience of CBT, treatment history)	2,552.17 (19, <.001)	320,080.02	.010	.017

Table 5.4 presents the odds ratios, with 95% confidence intervals, and Wald statistics for the independent variables within the regression model.

Table 5.4. Logistic regression results for predicting commencement of Module 1 (n=340,313).

	Odds Ratio	**95% CI**	***p***	**Wald**
Gender				
Female	**1.08*****	[1.06, 1.10]	<.001	69.66
Age group * *reference group: ≥ 50 years*				
≤ 19 years	**1.44*****	[1.38, 1.49]	<.001	351.67
20 – 29 years	**1.45*****	[1.41, 1.49]	<.001	681.89
30 – 39 years	**1.33*****	[1.29, 1.37]	<.001	406.14
40 – 49 years	**1.22*****	[1.19, 1.26]	<.001	188.91
Country * *reference group: Australia*				
United Kingdom	**1.10*****	[1.07, 1.12]	<.001	69.72
United States	**0.95****	[0.91, 0.98]	.005	7.80
Canada	**0.91*****	[0.88, 0.94]	<.001	29.65
Location				
Rural/remote	**1.10*****	[1.07, 1.12]	<.001	75.48

Table 5.4. Logistic regression results for predicting commencement of Module 1 (n=340,313) (continued).

	Odds Ratio	95% CI	*p*	Wald
Education level ** reference group: Higher degree*				
Some secondary	**0.74*****	[0.71, 0.78]	<.001	194.13
Secondary only	**0.92*****	[0.89, 0.96]	<.001	15.91
Trade /certificate	**0.87*****	[0.84, 0.89]	<.001	84.91
Undergraduate	1.04	[1.00, 1.09]	.070	3.29
Diploma	**0.87*****	[0.84, 0.91]	<.001	56.33
Bachelor's degree	0.99	[0.96, 1.02]	.446	0.58
Health professional referral				
Referred	**1.09*****	[1.07, 1.11]	<.001	75.01
Depression history				
Previous depression	**1.15*****	[1.12, 1.18]	<.001	112.24
Motivation level				
Motivated	**1.18*****	[1.15, 1.20]	<.001	221.22
Experience of CBT ** reference group: Not sure what CBT is*				
Tried, useful	0.99	[0.96, 1.02]	.521	0.41
Tried, not useful	**1.06****	[1.02, 1.11]	.006	7.52
Not tried, want to	**1.30*****	[1.27, 1.33]	<.001	521.39
Not tried, don't want to	**0.83*****	[0.79, 0.87]	<.001	64.30
Treatment history				
Number of treatments	**0.99*****	[0.98, 0.99]	<.001	120.23

$^{**}p < .01$, $^{***}p < .001$

After controlling for other registration variables, most variables were significant predictors of Module 1 commencement. Users were more likely to continue to the first module if they were female (8% greater odds), were from a rural or remote region (10% more likely), were referred by a health professional (9% more likely) or reported having previous depression (15% greater odds). Users had 18% greater odds of commencing Module 1 if they indicated at registration that they were motivated to complete the program.

Age group was the most important predictor in the model (Wald statistics from 188.91 to 681.89), with users aged over 50 having less odds of commencing the first module than other users (44% less than 19 years or younger; 45% less than 20-29 years; 33% less than 30-39 years; and 22% less than

40-49 years). Previous experience of CBT was also an important predictor. Compared to users who indicated that they were unsure what CBT was, users who had not previously experienced CBT but wanted to try it were 30% more likely to commence Module 1 (Wald: 521.39). There was no significant difference in the likelihood of starting Module 1 between users who had previously found CBT useful and those who were not sure what the therapy was.

Users from the U.K. were 10% more likely to commence Module 1 than Australian users, and Canadian and U.S. users were 9% and 5% less likely respectively. Lower education was also associated with less likelihood of a registrant commencing Module 1. Compared to users with a higher degree, users who had not completed secondary education had 26% less odds, users who had secondary schooling 8% less odds, users with a trade, certificate or apprenticeship 13% less odds, and users with a diploma 13% less odds. There were no significant differences in Module 1 commencement between users with a higher degree, and those who were current undergraduates or Bachelor's degree graduates.

Number of modules completed

Of the total number of users who commenced Module 1 (*n* = 339,756), 56.0% (190,289) completed this first module, but only a small number of users (2.9%) completed the entire program. Table 5.5 shows the number of modules these users completed.

Table 5.5. Number of modules completed by users who commenced Module 1 (n=339,759).

Number of modules	All users *n* (column %)	Top four country users *n* (column%)
0	149,467 (44.0%)	121,433 (43.7%)
1	132,397 (39.0%)	109,421 (39.3%)
2	39,372 (11.6%)	32,226 (11.6%)
3	6,687 (2.0%)	5,497 (2.0%)
4	1,961 (0.6%)	1,599 (0.6%)
5	9,872 (2.9%)	7,975 (2.9%)
Total	**339,756 (100%)**	**278,151 (100%)**

The characteristics of users from a top four country, who completed zero, one, or two or more modules are summarised in Table 5.6. Based on univariate analyses, completion of more modules was significantly associated with being motivated to complete the program at the time of registration, being referred by a health professional, and having no history of depression. Users were also significantly more likely to complete more modules if they reported that they had not previously

tried CBT but wished to try it. There was a significant association between the number of completed modules for different age groups. Younger users were significantly more likely to complete more modules; however, there was no significant difference in the number of modules completed between the 40-49 years and over 50 years age groups. Users from the United Kingdom were significantly more likely to complete more modules than users from each of the other three countries, and users with a trade, diploma or university degree were more likely to complete more modules than other users. Being located in a rural area was not significantly associated with the number of modules completed.

Users who completed one module had significantly higher initial depression and anxiety assessments scores and significantly higher dysfunctional thinking scores than users who completed either no modules or more than two modules. In addition, users who did not complete a module reported previously trying a higher number of evidence-based treatments for depression than users who completed one or more modules.

Table 5.6. Characteristics of top four country users who completed zero, one or more than two modules (n=278,151).

	Number modules completed			Chi-square
	0 modules ***n*** **(column %)**	**1 module** ***n*** **(column %)**	**2-5 modules** ***n*** **(column %)**	
Total	**121,443 (100%)**	**109,421 (100%)**	**47,297 (100%)**	
Gender				**17.95*****
Female	84,223 (69.4%)	76,773 (70.2%)	32,992 (69.8%)	
Male	37,220 (30.6%)	32,648 (29.8%)	14,305 (30.2%)	
Age group				**1089.02*****
≤ 19 years[a]	11,373 (9.4%)	13,519 (12.4%)	5,539 (11.7%)	
20 – 29 years[b]	36,373 (30.0%)	35,941 (32.8%)	14,868 (31.4%)	
30 – 39 years[c]	31,279 (25.8%)	26,816 (24.5%)	11,541 (24.4%)	
40 – 49 years[d]	24,279 (20.0%)	19,255 (17.6%)	8,715 (18.4%)	
≥ 50 years[d]	18,139 (14.9%)	13,890 (12.7%)	6,634 (14.0%)	
Country				**219.52*****
United Kingdom[a]	73,372 (60.4%)	68,189 (62.3%)	29,827 (63.1%)	
Canada[b]	9,879 (8.1%)	7,990 (7.3%)	3,553 (7.5%)	
United States[c]	7,513 (6.2%)	5,788 (5.3%)	2,554 (5.4%)	
Australia[c]	30,679 (25.3%)	27,454 (25.1%)	11,363 (24.0%)	
Location				3.04
Rural / remote	31,835 (26.2%)	29,000 (26.5%)	12,542 (26.5%)	
Urban	89,608 (73.8%)	80,421 (73.5%)	34,755 (73.5%)	
Education level				**716.39*****
Some secondary[a]	9,057 (7.5%)	9,368 (8.6%)	3,798 (8.0%)	
Secondary only[b]	11,679 (9.6%)	11,491 (10.5%)	4,349 (9.2%)	
Trade /certificate[c]	28,723 (23.7%)	26,383 (24.1%)	10,174 (21.5%)	
Undergraduate[a,b]	9,372 (7.7%)	9,596 (8.8%)	3,764 (8.0%)	
Diploma[d]	14,980 (12.3%)	12,846 (11.7%)	5,313 (11.2%)	
Bachelor's degree[c]	30,978 (25.5%)	26,359 (24.1%)	12,801 (27.1%)	
Higher degree[c,d]	16,654 (13.7%)	13,378 (12.2%)	7,098 (15.0%)	
Health professional referral				**390.23*****
Referred	56,498 (46.5%)	54,675 (50.0%)	24,065 (50.9%)	
Not referred	64,945 (53.5%)	54,746 (50.0%)	23,232 (49.1%)	
History of depression				**69.62*****
History	102,939 (84.8%)	92,808 (84.8%)	39,383 (83.3%)	
No history	18,504 (15.2%)	16,613 (15.2%)	7,914 (16.7%)	

Table 5.6. Characteristics of top four country users who completed zero, one or more than two modules (n=278,151) (continued).

	Number modules completed			Chi-square
	0 modules *n* (column %)	**1 module *n* (column %)**	**2-5 modules *n* (column %)**	
Motivation level				**1725.39*** **
Motivated	27,117 (22.3%)	27,261 (24.9%)	15,171 (32.1%)	
Not motivated	94,326 (77.7%)	82,160 (75.1%)	32,126 (67.9%)	
Experience of CBT				**783.52*** **
Tried, useful[a]	17,690 (14.7%)	13,759 (12.7%)	6,853 (14.6%)	
Tried, not useful[a]	7,776 (6.5%)	6,340 (5.9%)	2,720 (5.8%)	
Not tried, want to[b]	63,025 (52.5%)	58,794 (54.3%)	26,978 (57.7%)	
Not tried, don't want to[a.c]	4,058 (3.4%)	3,564 (3.3%)	1,349 (2.9%)	
Not sure what it is[c]	27,667 (23.0%)	25,868 (23.9%)	8,891 (19.0%)	
	0 modules *M* (*SD*)	**1 module *M* (*SD*)**	**2-5 modules *M* (*SD*)**	**Welch's *F* (df1, df2)**
Goldberg Depression (baseline)	6.31 (1.86)[a]	6.41 (1.80)[b]	6.32 (1.90)[a]	**(2, 128,414.8) = 94.41*** **
Goldberg Anxiety (baseline)	6.45 (1.96)[a]	6.55 (1.90)[b]	6.49 (1.94)[c]	**(2, 128,894.9) = 80.01*** **
Warpy Thoughts (baseline)	46.28 (11.51)[a]	46.76 (11.39)[b]	46.60 (11.70)[c]	**(2, 128,928.5) = 52.39*** **
Number of treatments previously tried	3.28 (2.29)[a]	3.12 (2.21)[b]	3.24 (2.24)[c]	**(2, 130,189.5) = 152.40*** **

***$p < .001$

For categorical variables with more than two categories, each superscript letter denotes a subset of variable categories whose column proportions do not significantly differ from each other (after correction for multiple comparisons).

For continuous variables, each superscript letter denotes a subset of the dependent variable categories whose mean differences do not significantly differ from each other (Games-Howell post hoc tests).

A multinomial logistic regression analysis was applied to the data to investigate which registration or initial assessment variables predicted how many modules users would complete. The model was run with all independent variables; however, likelihood ratio tests showed that the 'rural location' variable did not significantly contribute to the model. Accordingly, the model was rerun without this variable. Variance Inflation Factors were calculated for each pairwise combination of independent variables indicating that multicollinearity was not problematic in the regression model.

The final model was statistically significant χ^2 (50) = 4,631.87, $p < .001$). The Nagelkerke pseudo R^2 was 0.019 which showed that the independent variables in the model only accounted for 1.9% of the variance in whether users completed zero, one or two or more modules (see Table 5.7).

Table 5.7. Multinomial logistic regression model test statistics.

	Chi-square (*df*, *p*)	**-2 Log likelihood**	**Cox & Snell R^2**	**Nagelkerke R^2**
Model (gender, age, country, education, referral, previous depression, motivated, experience of CBT, baseline depression, baseline anxiety, baseline Warpy Thoughts, number of treatments)	4,631.87 (50, <.001)	560,471.88	.017	.019

Table 5.8 presents the odds ratios, with 95% confidence intervals, and Wald statistics for the independent variables within the regression model.

Table 5.8. Multinomial regression results for predicting completion of one module compared to zero modules, and two or modules compared to zero modules (n=278,151).

	One module vs zero modules				2-5 modules vs. zero modules			
	Odds Ratio	**95% CI**	***p***	**Wald**	**Odds Ratio**	**95% CI**	***p***	**Wald**
Gender								
Female	**1.03****	[1.01, 1.05]	.004	8.32	1.01	[0.98, 1.03]	.494	0.47
Age group* *reference group: ≥ 50 years*								
≤ 19 years	**1.45*****	[1.39, 1.51]	<.001	374.64	**1.47*****	[1.40, 1.54]	<.001	233.01
20 – 29 years	**1.25*****	[1.22, 1.29]	<.001	244.82	**1.08*****	[1.04, 1.12]	<.001	19.13
30 – 39 years	**1.10*****	[1.07, 1.13]	<.001	47.15	**0.96***	[0.92, 0.99]	.018	5.58
40 – 49 years	1.02	[0.99, 1.05]	.165	1.93	**0.96***	[0.93, 1.00]	.042	4.14
Country * *reference group: Australia*								
United Kingdom	1.00	[0.98, 1.02]	.890	0.02	**1.03***	[1.01, 1.06]	.016	5.83
United States	**0.86*****	[0.83, 0.89]	<.001	59.33	**0.90*****	[0.85, 0.94]	<.001	17.23
Canada	**0.90*****	[0.87, 0.93]	<.001	38.09	**0.93****	[0.89, 0.97]	.001	10.56
Education level * *reference group: Higher degree*								
Some secondary	**1.09*****	[1.05, 1.14]	<.001	17.68	**0.88*****	[0.83, 0.93]	<.001	22.81
Secondary only	**1.08*****	[1.05, 1.12]	<.001	19.35	**0.81*****	[0.77, 0.84]	<.001	82.60
Trade /certificate	**1.07*****	[1.04, 1.10]	<.001	18.43	**0.80*****	[0.77, 0.83]	<.001	143.51
Undergraduate	**1.09*****	[1.05, 1.13]	<.001	17.81	**0.85*****	[0.80, 0.89]	<.001	43.50
Diploma	1.02	[0.98, 1.05]	.306	1.05	**0.80*****	[0.77, 0.84]	<.001	103.77
Bachelor's degree	1.02	[0.99, 1.05]	.131	2.28	**0.95***	[0.92, 0.99]	.010	6.67

Table 5.8. Multinomial regression results for predicting completion of one module compared to zero modules, and two or modules compared to zero modules (n=278,151) (continued).

	One module vs zero modules				2-5 modules vs. zero modules			
	Odds Ratio	**95% CI**	***p***	**Wald**	**Odds Ratio**	**95% CI**	***p***	**Wald**
Health professional referral								
Referred	**1.11*****	[1.09, 1.13]	<.001	142.60	**1.21*****	[1.18, 1.23]	<.001	258.57
Depression history								
Previous depression	1.01	[0.98, 1.03]	.503	0.45	**0.89*****	[0.86, 0.92]	<.001	48.98
Motivation level								
Motivated	**1.16*****	[1.14, 1.19]	<.001	224.96	**1.61*****	[1.57, 1.65]	<.001	1502.74
Experience of CBT ** reference group: Not sure what it is*								
Tried, useful	**0.94*****	[0.91, 0.97]	<.001	14.31	**1.18*****	[1.13, 1.23]	<.001	55.51
Tried, not useful	**0.95***	[0.91, 0.99]	.013	6.10	**1.11*****	[1.05, 1.17]	<.001	13.31
Not tried, want to	**1.05*****	[1.03, 1.07]	<.001	18.78	**1.30*****	[1.27, 1.34]	<.001	311.01
Not tried, don't want to	**0.95***	[0.91, 1.00]	.040	4.30	1.00	[0.94, 1.07]	.986	0.00
Goldberg Depression (baseline)	**1.02*****	[1.02, 1.03]	<.001	67.50	1.00	[1.00, 1.01]	.476	0.51
Goldberg Anxiety (baseline)	**1.01*****	[1.01, 1.02]	<.001	20.26	**1.01***	[1.001, 1.01]	.030	4.70
Warpy Thoughts (baseline)	1.00	[1.00, 1.00]	.609	0.26	1.00	[1.00, 1.00]	.872	0.03
No. of previous treatments	**0.98*****	[0.98, 0.99]	<.001	48.19	1.00	[0.99, 1.01]	.816	0.05

*p < .05, **p < .01, ***p < .001

The most important predictor in the regression model was whether users indicated at registration that they were motivated to complete the program. After controlling for the other variables, those who were motivated were 16% more likely to complete one module rather than no modules, and 61% more likely to complete two or more modules compared to no modules. In addition, users who endorsed that they had not tried CBT but wanted to try it were 5% more likely to complete one compared to no modules than users who were not sure what CBT was, and 30% more likely to complete two or more modules compared to no modules. Users who were referred to the site by a health professional also had higher odds of completing more modules (11% greater odds for completing one module and 21% greater odds of completing two or more modules).

Controlling for other factors, younger users had greater odds of completing more modules – compared to users over 50 years old, users aged less than 19 years were 45% and 47% more likely to complete one module or two or more modules, and users aged 20 to 29 years were 25% and 8% more likely. The contribution of education level was mixed, with users who were less educated more likely to complete one module but less likely to complete two or more modules. Users from Australia had greater odds of completing more modules than users from Canada or the United States; however, users from the United Kingdom were more likely than Australian users to complete two or more modules.

Having a self-reported history of depression was not a significant predictor of completion of one module, but users with a depression history were 11% less likely to complete two or more modules. Conversely, each additional depression treatment a user had previously tried corresponded to a 2% decrease in odds of completing one module, but the number of treatments did not significantly predict completion of two or more modules.

The initial symptom scores only contributed a small amount to the model. A one-unit increase on the baseline Goldberg depression score was associated with a 2% increase in odds of completing one module but was not significantly associated with completing two or more modules. A one-unit increase on the Goldberg anxiety score was associated with a 1% increase in odds of completing either one module or two or more modules. Warpy Thoughts score was not a significant predictor of the number of modules completed.

Number of logins

On average, users who commenced Module 1 logged in to the program 2.54 times (n = 339,756, SD = 3.13), with nearly half of these users accessing the program only once (n = 164,235, 48.3%). The number of logins ranged up to 387 times; however, nearly all participants (97.1%) logged in fewer

than 10 times. The number of logins is displayed in Figure 5.2. This scatterplot does not include data for 11 users who had more than 100 logins.

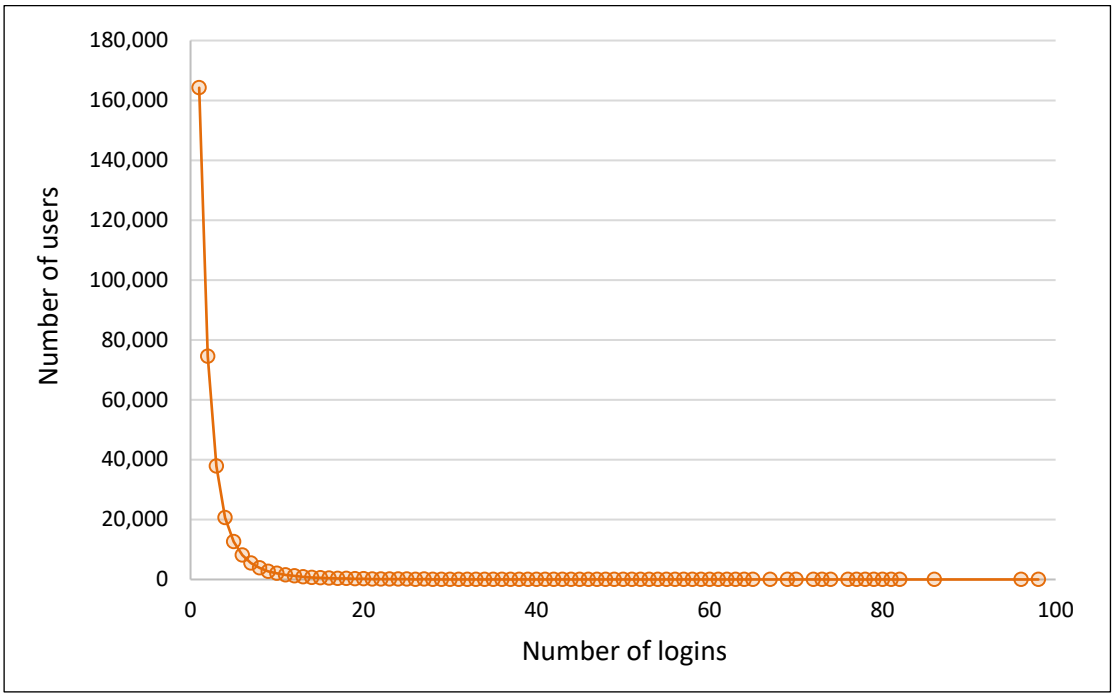

Figure 5.2. Number of logins for users who commenced Module 1 (n=339,745). The graph omits the data of 11 users with more than 100 logins.

The numbers of users who logged in for each number of times up to 10, and for 10 or more times, is displayed in Table 5.9.

Table 5.9. Number of logins for users who commenced Module 1 (n=339,756).

No. logins	All users n (column %)	Top four country users n (column%)
1	164,235 (48.3%)	134,913 (48.5%)
2	74,508 (21.9%)	61,336 (22.1%)
3	37,828 (11.1%)	30,996 (11.1%)
4	20,679 (6.1%)	16,931 (6.1%)
5	12,596 (3.7%)	10,303 (3.7%)
6	8,178 (2. 4%)	6,625 (2.4%)
7	5,438 (1.6%)	4,422 (1.6%)
8	3,858 (1.1%)	3,106 (1.1%)
9	2,710 (0.8%)	2,131 (0.8%)
≥ 10	9,726 (2.9%)	7,388 (2.7%)
Total	**339,756 (100%)**	**278,151 (100%)**

Time on the program

Users spent an average of 79.6 minutes in total on the program module webpages (*n* = 339,756, *SD* = 108.28), ranging from under 1 minute to 10,150 minutes (169.2 hours). The number of minutes (rounded to whole minutes) spent on the site by users is displayed in Figure 5.3, excluding data from nine users who spent more than 3,000 minutes on the program.

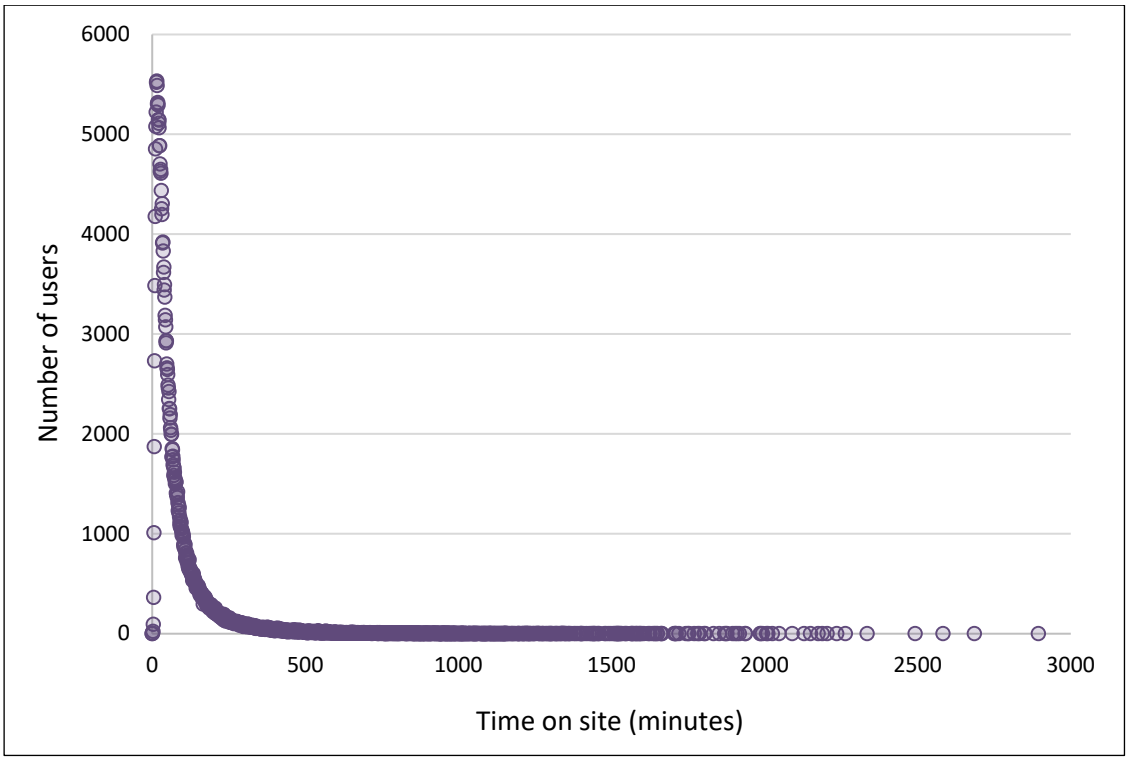

Figure 5.3. Time on the program (rounded minutes) for users who commenced Module 1 (n=339,756). The graph omits the data for 9 users with more than 3,000 minutes on the program.

The number of users who spent less than 30 minutes, between 30 and 60 minutes, between 60 and 120 minutes, and 120 minutes or more on the program is displayed in Table 5.10 below.

Table 5.10. Number of minutes spent on the program, for users who commenced Module 1 (n=339,756).

Time on program	All users n (column %)	Top four country users n (column%)
< 30 mins	112,452 (33.1%)	92,735 (33.3%)
30-60 mins	92,584 (27.3%)	76,378 (27.5%)
60-120 mins	73,907 (21.8%)	60,474 (21.7%)
≥ 120 mins	60,813 (17.9%)	48,564 (17.5%)
Total	**339,756 (100%)**	**278,151 (100%)**

5.3.2 Aim 2: Depression symptom changes

Of the 339,756 users who commenced the first module, 130,426 (38.4%) completed at least one more depression assessment survey at the start of Module 2 or later in the program. Similarly, 106,882 (38.4%) of the 278,151 users from one of the top four countries completed at least two depression surveys.

A paired-samples *t*-test showed a statistically significant decrease in Goldberg depression score for users from any country, from initial assessment (M = 6.34, SD = 2.24) to post-test [M = 5.76, SD = 1.86, $t(130{,}425) = 117.50$, $p < .001$], with a small effect size (Cohen's d = 0.28). The mean change score was 0.58 (95% CI: 0.57, 0.59), indicating a significant decrease in the overall number of depression symptoms reported by users.

Effect of adherence and related variables on depression symptom score changes

Univariate analyses were conducted to investigate the impact of adherence to the intervention on depression symptom changes, after controlling for initial depression score. A greater improvement in a user's depression score was significantly associated with higher adherence to the intervention, when measured by the number of modules completed, the number of logins and time spent on the program, as described below. In addition, there was a greater improvement in depression symptoms where a longer time had elapsed between initial and post-test assessments. These analyses are reported for users from the top four countries since this sub-group of users is the focus of later multivariate regression analyses (see next section, pp. 197-200).

Number of modules completed

The change in depression score was significantly correlated with the number of modules completed (r = 0.163, n = 106,882, $p < .001$). Further, there was a significant effect of the number of modules completed on the depression change scores after adjustment for initial depression severity [$F(4, 106{,}876) = 904.57$, $p < .001$]. Figure 5.4 shows the estimated marginal means according to the number of modules completed. When adjusted for multiple comparisons, pairwise comparisons between different numbers of completed modules were all significant at $p < .001$ or, for the contrast between three and four modules, at $p = .005$.

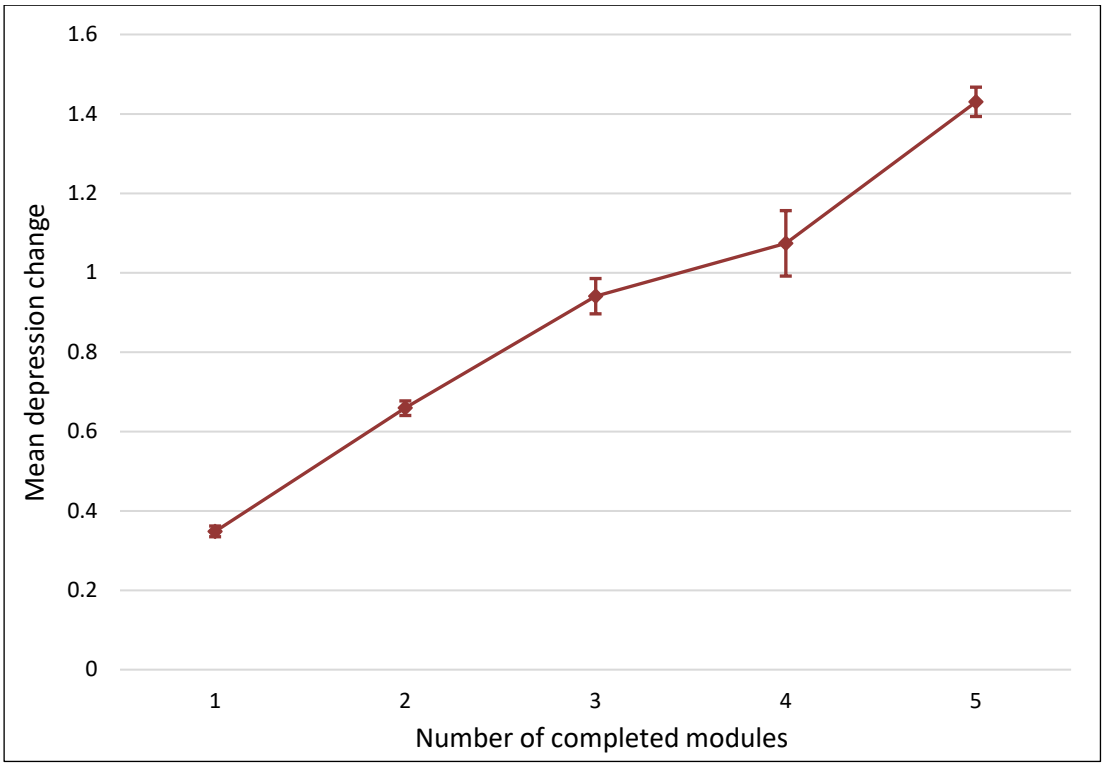

Figure 5.4. Estimated mean changes in depression score by number of modules completed for users from the top four countries (n=106,882).
Error bars represent 95% confidence intervals. Initial depression score set at mean value (6.39).

Number of logins

The number of times a user accessed the program was also positively correlated with depression symptom change (r = 0.167, n = 106,882, p < .001). After controlling for initial depression score, change in depression scores ranged from 0.21 (95% CI: 0.18, 0.23) for users who completed only a single session, to 1.33 (95% CI: 1.29, 1.36) for users who logged in ten or more times. There was a significant effect of the number of sessions on depression change scores [F(9, 106,871) = 1,143.62, p < .001]. The estimated marginal means are displayed in Figure 5.5. After adjusting for multiple comparisons, most differences in depression change scores for different numbers of logins were significant, except for the differences between 6 and 7, and 8 and 9 logins.

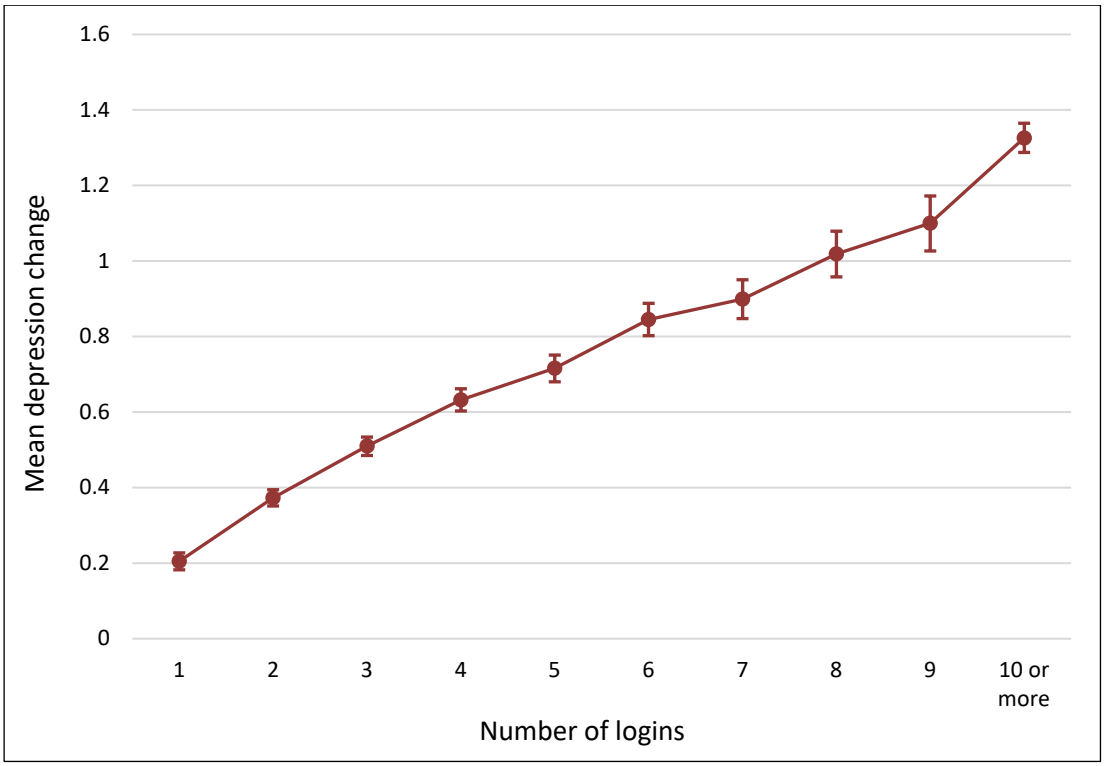

Figure 5.5. Estimated mean changes in depression score by number of logins for users from the top four countries (n=106,882).

Error bars represent 95% confidence intervals. Initial depression score set at mean value (6.39).

Time spent on the program

Similarly, there were significantly greater changes in depression for users who spent more time on the site [$F(3, 106{,}877) = 2{,}537.29$, $p < .001$]. The depression change score for users who spent more than two hours on the program was estimated to be 1.43 (95% CI: 1.39, 1.47) after adjustment for initial depression score, whereas the change for users who spent less than 30 minutes was 0.35 (95% CI: 0.33, 0.36). Figure 5.6 shows the estimated marginal means for change in depression scores across the different levels of time spent on the program. Pairwise comparisons were highly significant ($p < .001$) between each combination of levels, after adjustment for multiple comparisons.

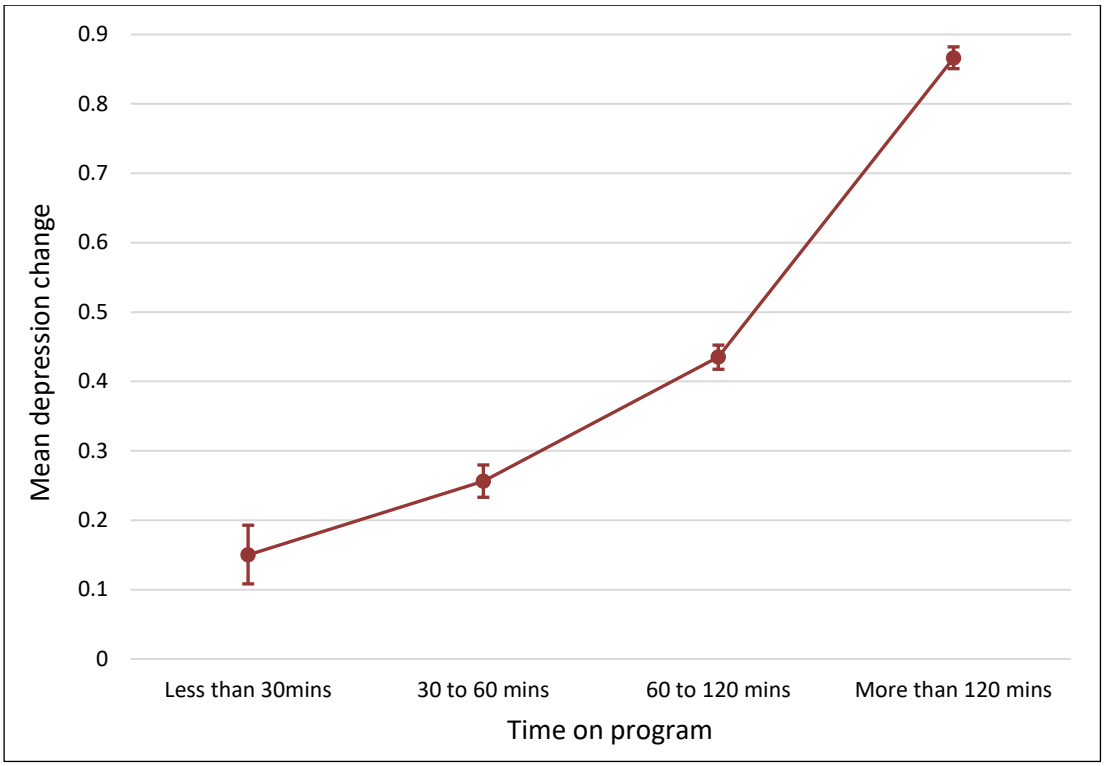

Figure 5.6. Estimated mean changes in depression score by time spent on the program for users from the top four countries (n = 106,882).

Error bars represent 95% confidence intervals. Initial depression score set at mean value (6.39).

Time elapsed between pre- and post-tests

After controlling for initial depression score, there was also a significant effect of time between tests on the depression change scores [$F(2, 106{,}878) = 4{,}301.28$, $p < .001$]. The estimated marginal mean change score for users who completed tests more than 1 week apart was 0.89 (95% CI: 0.87, 0.91), whereas the estimated mean for users who completed the tests on the same day was 0.24 (95% CI: 0.23, 0.26), as shown in Figure 5.7. All six pairwise comparisons were highly significant after applying a correction for multiple comparisons ($p < .001$).

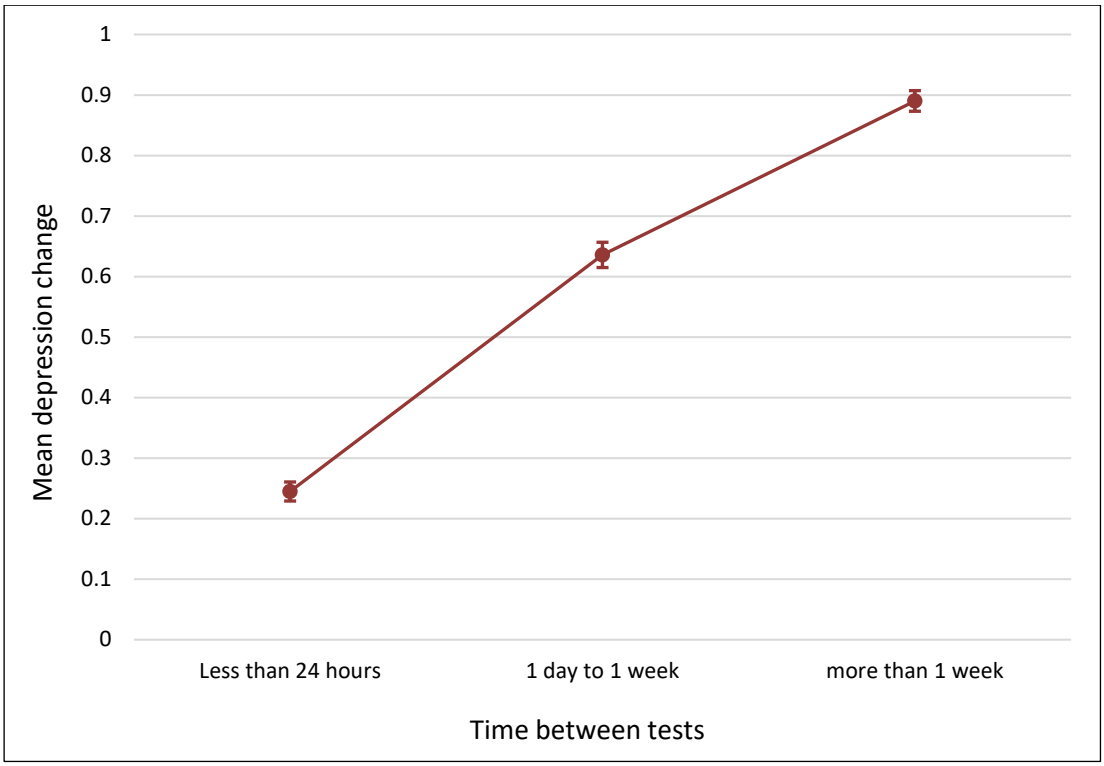

Figure 5.7. Estimated mean changes in depression score by time between pre-test and post-test for users from the top four countries (n=106,882).
Error bars represent 95% confidence intervals. Initial depression score set at mean value (6.39).

Correlations between adherence variables

The Pearson Correlation coefficients for each pair of adherence measures were highly significant ($p < .001$), and the number of modules completed, the number of logins and time on the program were strongly associated with each other (see Table 5.11). In addition, the time elapsed between tests was strongly correlated with the number of logins ($r = 0.58$) and time spent on the program ($r = 0.413$).

Table 5.11. Pearson Correlation coefficients (r) for pairs of adherence measures (n=106,882). All associations are significant (p<.001).

	No. modules	**No. logins**	**Time on program**	**Time between tests**
No. modules	1	0.435	0.420	0.269
No. logins	0.435	1	0.559	0.584
Time on program	0.420	0.559	1	0.413
Time between tests	0.269	0.584	0.413	1

Predicting depression symptom changes

After examining the univariate effect of each of the independent variables on change in depression score (controlling for initial depression score), a hierarchical multivariate regression analysis was undertaken. Initial depression, anxiety and dysfunctional thinking scores were entered in the first block, followed by demographic and other registration characteristics in the second block. Independent variables were retained in this block if they made a statistically significant contribution to the model, with the block including age group, education level, country, rural location, history of depression, motivation level, experience of CBT, number of treatments tried, and previous use of anti-depressant medication. Referral to the program by a health professional and previous use of CBT were not significant after controlling for other variables and were not included in the model. In the final block, 'process' or adherence variables were added. Given the correlation between adherence variables, only the number of modules and time elapsed between pre-test and post-test were included.

All three models were statistically significant, as shown in Table 5.12. The model including initial depression, anxiety and dysfunctional thinking scores, accounted for 8.9% of the variance in the depression change score. The addition of registration variables only explained an additional 1.6% of the variance, whereas the inclusion of adherence variables in the final step explained a further 3.9% or the variance. Variance Inflation Factors for the independent variables did not exceed 2.7 in any of the three models, suggesting that multicollinearity was not an issue in the regressions. The Beta coefficients for the independent variables in each of the three models are displayed in Table 5.13.

Table 5.12. Multivariate regression model test statistics for depression change score.

	R^2	R^2 change	*F* change	Sig *F* change
Step 1 (baseline depression, anxiety, dysfunctional thinking)	0.089	0.089	3,459.10	<.001
Step 2 (Step 1 + gender, age, education, country, location, motivated, depression history, experience of CBT, no. treatments, anti-depressants)	0.104	0.016	81.86	<.001
Step 3 (Step 2 + no. modules + time between tests)	0.144	0.039	1,635.16	<.001

Table 5.13. Hierarchical multivariate regression predicting depression change score (n=106,882).

	Step 1		Step 2		Step 3	
	β	*P*	β	*p*	β	*p*
Depression (baseline)	**0.318*****	<.001	**0.345*****	<.001	**0.354*****	<.001
Anxiety (baseline)	**-0.147*****	<.001	**-0.139*****	<.001	**-0.135*****	<.001
Warpy Thoughts (baseline)	**-0.092*****	<.001	**-0.088*****	<.001	**-0.090*****	<.001
Female			**0.008****	.006	**0.011*****	<.001
Age group ** reference group: ≥ 50 years*						
≤ 19 years			**-0.020*****	<.001	**-0.012****	.003
20 – 29 years			-0.005	.249	0.008	.075
30 – 39 years			-0.001	.803	**0.010***	.023
40 – 49 years			-0.008	.057	-0.002	.558
Country ** reference group: Australia*						
United Kingdom			**-0.025*****	<.001	**-0.030*****	<.001
United States			0.006	.058	**0.007***	.025
Canada			0.005	.159	0.001	.678
Education level ** reference group: Higher degree*						
Some secondary			**-0.039*****	<.001	**-0.031*****	<.001
Secondary only			**-0.035*****	<.001	**-0.025*****	<.001
Trade /certificate			**-0.042*****	<.001	**-0.028*****	<.001
Undergraduate			**-0.018*****	<.001	**-0.013*****	<.001
Diploma			**-0.019*****	<.001	**-0.009***	.021
Bachelor's degree			-0.003	.456	-0.002	.673
Rural/remote location			**0.009****	.004	**0.007***	.019
Depression history			**-0.065*****	<.001	**-0.063*****	<.001
Motivated			**0.059*****	<.001	**0.048*****	<.001
Experience of CBT ** reference group: Not sure what it is*						
Tried, useful			**0.035*****	<.001	**0.026*****	<.001
Tried, not useful			**-0.011****	.001	**-0.013*****	<.001
Not tried, want to			**0.036*****	<.001	**0.026*****	<.001
Not tried, don't want to			-0.005	.081	-0.004	.145
No. treatments			**0.017*****	<.001	0.005	.254
Anti-depressant use			**-0.038*****	<.001	**-0.031*****	<.001
No. of modules					**0.134*****	<.001
Time between tests ** reference group: More than one week*						
< 24 hours					**-0.131*****	<.001
1 day to 1 week					**-0.039*****	<.001

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Higher initial depression assessment scores significantly predicted larger positive change in depression symptoms and was the strongest predictor in all models [Step 3: B = .338, β = .354, t = 109.19, p < .001]. By contrast, higher initial anxiety and higher dysfunctional thinking were strong predictors of *reduced* depression improvement in all models [Anxiety: Step 3: B = -.124, β = -.135, t = -42.96, p < .001; Warpy Thoughts: Step 3: B = -.014, β = -.090, t = -29.32, p < .001].

The adherence and process variables were also important contributors to the final regression model. Completing more modules predicted greater depression improvement, with an additional 0.21-point improvement on the Goldberg depression scale predicted by each additional module completed [Step 3: B = .209, β = .134, t = 45.36, p < .001]. The time between the pre- and post-tests was also important. Compared to users who completed the tests over 1 week apart, users who completed both tests within 24 hours, or between 1 day and 1 week, had less depression symptom improvement [Less than 24 hours: Step 3: B = -.469, β = -.131, t = -39.01, p < .001; Between 1 day and 1 week: Step 3: B = -.162, β = -.039, t = -12.07, p < .001].

Of the registration characteristics of users, the most important variables were self-reported history of depression and whether the user was motivated to complete the program. Depression history was a significant predictor of reduced depression improvement in both the Step 2 and Step 3 models, [Step 3: B = -.303, β = -.063, t = -19.82, p < .001], whereas self-reported motivation to complete the program resulted in increased improvement in depression symptoms in both models [Step 3: B = .187, β = .048, t = 16.75, p < .001]. An increase in the number of depression treatments previously tried by a user was associated with increased depression improvement in the Step 2 model, but this variable was no longer significant after controlling for adherence variables in the final model. Conversely, self-reported use of anti-depressant medication was a predictor of less improvement in both models [Step 3: B = -.108, β = -.031, t = -7.93, p < .001]. Having previously tried CBT and finding it useful, or not having tried CBT but wanting to try it both predicted increased depression improvement in users [Tried, useful: Step 3: B = .134, β = .026, t = 6.63, p < .001; Not tried, want to: Step 3: B = .092, β = .026, t = 6.94, p < .001], whereas previously finding CBT not useful predicted decreased depression improvement after controlling for other variables [Tried, not useful: Step 3: B = -.102, β = -.013, t = 3.99, p < .001].

Compared to being located in Australia, location in the United Kingdom predicted less depression improvement in users [Step 3: B = -.108, β = -.030, t = -8.65, p < .001] whereas United States users demonstrated slightly more improvement [Step 3: B = .054, β = .007, t = 2.24, p < .001]. Logging in from Canada compared to Australia was not a significant predictor of depression change scores. Being located in a rural or remote region was a weak but significant predictor of increased

depression improvement, although it had a lower level of statistical significance compared to other variables (p = .019).

Users aged 19 years or younger were significantly less likely than users aged 50 years or older to improve on the Goldberg Depression scale in both the Step 2 model and the final model which controlled for adherence and related variables [Step 3: B = -.066, β = -.012, t = -2.92, p < .001]. Users with less educational attainment were also significantly less likely than those with greater educational attainment to demonstrate an improvement in depression symptoms in both models. In contrast, female gender predicted greater improvement in depression [Step 3: B = .043, β = .011, t = 3.87, p < .001].

5.4 Discussion

5.4.1 Summary of results

The results for the two aims of the study in this chapter are summarised below.

Aim 1: Program adherence and predictors

Of the 417,354 personal users who registered during the study period, 364,292 (87.3%) completed the initial depression assessment survey, and 81.4% (n = 339,756) subsequently commenced the first program module. On average users logged in 2.54 times and spent 79.6 minutes on the program webpages. The majority of participants who commenced Module 1 did not complete more than one module – 44% did not complete the first module and 39% completed only the first module. The remaining 17% completed two or more modules, with a small percentage (2.9%) completing all five modules.

Commencing Module 1 was predicted by: younger age; not having previously undertaken CBT but wishing to try it; self-reported motivation to complete the program; having a higher educational level; self-reported history of depression; location in a rural or remote region; referral to the program by a health professional; being in the United Kingdom; and being female.

Completing modules was best predicted by user motivation at the time of registration (61% more likely to complete two or more modules compared to no modules; 16% more likely to complete 1 vs 0 modules). Users who indicated that they had not tried CBT but wished to do so were also more likely to complete more modules (5% more likely to complete 1 vs 0 modules; 30% to complete 2 or more vs 0 modules). Referral to the program by a health professional was also an important predictor of progress (11% more likely to complete 1 vs 0 modules; 21% more likely to complete 2 or

more vs 0 modules). Progress through the program was also predicted by being younger and being from the United Kingdom. Users with a history of depression were less likely to complete two or more modules compared to no modules, but a history of depression did not predict completion of one module. Conversely, female users were more likely to complete one module compared to zero modules, but gender was not a significant predictor of completion of two or more modules. Higher educational attainment was associated with greater odds of completing two or more modules, but less odds of completing one module, compared to no modules. Completion of one compared to zero modules was predicted by higher initial depression assessment score and lower number of depression treatments previously tried, but neither variable predicted completion of two or more modules. Higher baseline anxiety score predicted completion of both one compared to zero modules, and two or more compared to zero modules.

Aim 2: Depression symptom changes and predictors

Two or more depression assessments were completed by 38.4% (*n* = 130,426) of users who commenced Module 1, which corresponded to 31.2% of all registered users. There was a significant, small decrease in mean depression symptom scores from initial assessment survey to the last completed survey for these users [mean score change = 0.58 (95% CI: 0.57, 0.59); Cohen's *d* = 0.28].

Higher adherence was associated with significantly better depression outcomes, with a greater reduction in depression scores positively associated with the number of modules completed, number of logins and time on the site. This association (represented by the number of modules completed) remained significant and was an important predictor of depression score change in a multivariate regression model controlling for registration and clinical characteristics. Increased time between pre- and post- tests also predicted a greater reduction in depression survey scores.

However, the most important predictor of improvement in depression status was initial depression score, with higher initial scores predicting greater improvement. The user registration variables which predicted a greater reduction in depression scores after controlling for adherence and clinical characteristics were: self-reported history of depression; being motivated to complete the program; having previously tried CBT and found it useful, or not having tried CBT but wanting to try it; being located in Australia (compared to the United Kingdom); being in a rural or remote region; and being female. Less depression improvement was predicted by being aged 19 years or younger, having a lower level of education, self-reported use of anti-depressant medication, higher dysfunctional thinking scores and, in contrast to the findings for depression severity, higher initial anxiety.

5.4.2 *User progress through the program*

The percentage of total users in the current study who completed at least one Goldberg depression assessment (89.2%) was much higher than for previous studies of spontaneous community users of *MoodGYM.* Only 61.9% of Version 1 users registered up until September 2003 and 51% of Version 2 users registered before October 2004 completed one or more depression assessments [132]. Similarly, in a larger study involving Version 2 users who registered over 16 months from January 2006, only 57.7% of those users who registered completed at least one depression assessment [133]. Although the depression assessments were not compulsory in Version 1 of *MoodGYM*, depression scale completion was a pre-requisite for progressing through Version 2 as it was for Version 3, the subject of the current study. In addition, the content of the program was consistent across versions. Given this similarity in the content and user experience for Versions 2 and 3, it is unclear why a larger proportion of users in the current study completed an assessment. It is possible that users in the current study, who registered later (from 2009), were more comfortable using technology given increased usage worldwide (including for health management) and were thus less likely to drop out of the program early. They may also have enjoyed superior Internet connectivity given the improvements in Internet infrastructure over time, decreasing the likelihood of immediate dropout. Further, given the growth of awareness about iCBT for depression, the more recent users may have had more realistic expectations about what was involved in iCBT.

Alternatively, it is possible that the improved rate of completion of assessments was associated with the change in the distribution of the characteristic of users over time. For example, although the age and gender distribution remained largely unchanged over time, a greater percentage of users in the current sample were referred by health professionals than those from the previous study of users who registered from January 2006 (48.2% referred in the current study; 17.8% referred in the previous study), and a greater percentage of users were from the U.K. (49.5% in the current study; 36.1% from Europe in the previous study) [133]. A regression analysis confirmed that health professional referral and being from the U.K. were both significant predictors of completing at least one depression test in the current sample (see Appendix B: Additional analyses, section B.1).

In contrast to the current study, previous studies of *MoodGYM* have not reported on predictors of *commencement of the program* after registration, although a study of 17,318 spontaneous, community users of the *HDep* program found that only 71.4% (*n* = 12,366) of users who registered on the program accessed the first module [135], which is comparable to the proportion of users who commenced *MoodGYM* in the current study (81.4%, *n* = 339,756). Significantly, users who indicated that they were motivated to complete the program, or who had not previously experienced CBT but

wished to try it, were more likely to commence Module 1 after registration and the initial assessment surveys. After controlling for other variables, being referred by a health professional also predicted commencement of Module 1, possibly because the user was undertaking the program based on the authority of a trusted professional and was, therefore, more committed to engaging in the recommended treatment or because such users had an expert-verified unmet need. Users with a lower educational level were less likely to commence the program, perhaps due to the level of reading and comprehension required by *MoodGYM* [170]. Possible explanations for other predictors are more tenuous. Users who failed to commence the program may have decided that the program would be unsuitable based on some aspect of the introduction such as the program overview, the character descriptions, the surveys, or the user interface. Perhaps, for example, the introductory content and style were considered less suitable by male users and older users. On the other hand, increased likelihood of the commencement of the program could reflect a lack of other potential depression treatment resources, such as could be the case for users from rural areas.

Further, factors that were not measured in the current study might mediate the relationship between demographic characteristics and the likelihood that a user will progress through the program. For example, Ritterband and colleagues identified other potentially relevant user characteristics for predicting behaviour change via Internet interventions, including personality traits, cognitive factors, readiness for change, physiological factors and computer skills [175]. In addition, the environment within which the user accessed the program may impact the user's acceptance and engagement with a self-help program. This potentially includes influences at a social level (such as family and friends), the setting within which the program was introduced (e.g., school, work, primary care), and at a societal level (e.g., cultural or media) [175].

Although two previous studies of self-guided spontaneous community users have investigated the predictors of the number of modules completed [133, 134], none have previously investigated the potential role of the user's motivation in completing the program, nor of the contribution of their previous experience of CBT. Significantly, these variables were important contributors to the prediction model, with users who were motivated, and those who had previously found CBT useful or not previously experienced CBT but wished to do so, more likely to complete more modules. This suggests that pre-intervention information about iCBT and motivational interviewing [176] designed to increase motivation may promote program adherence among users who subsequently choose to access the program.

After controlling for motivation and previous experience of CBT, many of the significant predictors of module completion in the current analysis were consistent with the results of the previous

MoodGYM studies (for example, health professional referral, younger age group and location in Europe/United Kingdom). Higher education levels also predicted completion of two or more modules but not one module, consistent with earlier results. As reported in previous *MoodGYM* studies, rural location was not a significant predictor. Similarly, as in the current study, women were more likely to complete the *HDep* program [135] than men (although the latter program was designed for women so this finding is not unexpected), as were those with self-reported disability from mental illness (which is likely to correlate with depression and anxiety symptoms).

However, there were some differences between the findings of the current analysis and those of previous *MoodGYM* studies. In the current analysis, after controlling for other variables, higher initial depression scores did not predict completion of two or more modules, and level of dysfunctional thinking was not a significant predictor of completing either one, or two or more modules compared to zero modules. There was, however, a difference in the nature of the dysfunctional thinking measure included in the current compared to the previous analysis [133]. The latter study of predictors of adherence employed the summed score of all 42 assessment items [133] – an unvalidated measure – whereas the current study was based on the 20-item measure comprising the higher order factor that emerged from formal psychometric analysis of the original scale items. Nevertheless, in order to investigate if the discrepancy in findings from the two studies might be attributable to the use of the different measures, the multiple regression in the current study was rerun using the total summed score of the Warpy Thoughts Quiz. It showed that 42-item Warpy Thoughts Quiz score was only a weak predictor of module completion (one vs zero modules: OR = 1.0005, p = .007; two or more vs zero modules: OR = 1.0006, p = .009; see Appendix B: Additional analyses, section B.2), whereas it was strong predictor in the model reported in the previous study (one vs zero modules: OR = 1.038, p = .027; two or more vs zero modules: OR = 1.131, p < .001). Further investigation of the role of dysfunctional thoughts in adherence is required. However, given that the current study employed the psychometrically validated measure of dysfunctional thoughts, there is little convincing evidence that dysfunctional thinking promotes adherence once an individual commences a program. As noted above, the second difference between the findings of the current and previous studies is that in the latter a history of depression was not a significant predictor of adherence whereas in the current study such a history was associated with 11% less odds of completing two modules predictor in previous studies. The reason for this discrepancy is unclear.

Overall, the current study confirms that younger, female users and those who are referred by a health professional are more likely to complete more of the *MoodGYM* program. However, the user's initial levels of depression, anxiety and dysfunctional thinking may be relatively unimportant in

predicting adherence to the program after controlling for motivation and the user's previous experience of (or attitude to) CBT. In contrast to the finding of the current and previous studies of *MoodGYM* that younger users were more likely to complete more modules, the study of *HDep* users found the opposite [135]. This could because the *MoodGYM* program content was originally designed for users aged 18 to 25 years, with examples relevant to this cohort, and so might be more engaging for younger users.

Since only three other studies have investigated predictors of adherence among spontaneous self-guided iCBT users, it is of interest to compare the current findings with those from studies of the predictors of treatment adherence among other self-guided users of depression Internet interventions. As noted in Chapter 2, Karyotaki et al. undertook a quantitative meta-analysis of such RCTs [37]. The authors concluded that non-adherence was predicted by being male, having a lower educational level, younger age and comorbid anxiety symptoms. Thus, the current results were consistent with the findings reported by Karyotaki with respect to gender and educational levels. However, in the current study, higher anxiety symptoms did not predict completion of two or more modules, and the findings for the effect of age on module completion were the converse of those reported in the meta-analysis although consistent with those reported for the *HDep* program. As noted above, it is possible that younger users complete more of *MoodGYM* because it was originally written for young users. Hence, the finding that adherence to *MoodGYM* is greater among young people may not be generalisable to other iCBT programs for depression.

Overall, very few users completed the whole program (*n* = 9,872, 2.9% of those who started Module 1). This high level of attrition is often cited as a disadvantage of automated iCBT [e.g., 177]. However, as noted in previous chapters, the provision of automated iCBT to large numbers of users is cost-effective, with minimal extra resources required for delivery to additional users. It is noteworthy that 17% of those users who started Module 1 completed at least two modules (*n* = 47,297). The first two modules of the program introduce fundamental concepts about CBT and have been shown in a previous trial to be the most important modules for reducing depression symptoms [118]. Therefore, given that the program was provided as a freely available open-access program, the provision of basic CBT training to nearly 50,000 users over the 5-year period represents a potentially significant public health contribution.

5.4.3 Changes in depression symptoms of users

Overall there was a small, statistically significant reduction in depression symptoms over time among spontaneous community *MoodGYM* users who completed two or more Goldberg Depression

assessment surveys. This result confirms the findings of older studies of *MoodGYM* [130, 131, 132]. Further research is required to determine if these findings generalise to the real-world use of other depression iCBT programs.

Consistent with previous studies of spontaneous community users of *MoodGYM*, the magnitude of improvement was positively associated with the number of modules completed [130, 131, 132]. There have been no previous studies of the predictors of outcomes for spontaneous users of other depression iCBT programs. However, the result is consistent with the findings of a broader systematic review of the impact of adherence on the effectiveness of Internet interventions which found that progress through a program predicted psychological outcomes in research trials [178]. In addition, the finding in the current study that greater improvement in depression symptoms was independently associated with a longer elapsed time between pre- and post-assessments after controlling for other factors including the number of modules completed was consistent with the results of an earlier study of spontaneous users of *MoodGYM* [130]. The association between outcome and time elapsed may be attributable in part to the longer time available to implement and practice the CBT strategies learnt during the course of the program. In addition, mild depression symptoms often improve over time without intervention [179]. The fact that the time elapsed remained a significant predictor of depression change scores after controlling for the number of modules completed excludes the possibility that the 'time elapsed' was a simple proxy for 'modules completed'.

The final multivariate model found that higher initial depression assessment scores predicted greater improvement in depression score, an association which had not been examined in previous studies of spontaneous *MoodGYM* users. This suggests that the *MoodGYM* program may be most effective for those with a high level of depression severity, although, it is also possible those with high scores were less influenced by floor effects. However, by contrast, a meta-analysis of RCTs of outcomes for users of unguided iCBT for depression found that only adherence measures and not participant characteristics, including initial depression severity, moderated depression outcomes [28]. There are a number of possible explanations for this discrepancy. For example, the attributes of the spontaneous users in the current study may differ from those of participants in randomised controlled trials and the results may have been confounded by other factors such as the use of different interventions across the studies. Further research is required. Meanwhile, based on the current findings, we conclude that initial depression severity is associated with better outcomes for *MoodGYM* in particular, but that further research on individual interventions is required to determine if this relationship can be generalised to other iCBT programs for depression.

In contrast to the outcomes for users with high initial depression severity, higher initial anxiety scores predicted less depression improvement. This was the case in all models, including those which did not control for adherence. It is unclear why self-directed CBT for depression is less effective for those with a high level of comorbid anxiety symptoms given that iCBT is a known effective intervention for anxiety symptoms [180] and *MoodGYM* has been shown to reduce anxiety symptoms in those with elevated depressive symptoms [181]. However, it is possible that such users would benefit from an approach that involves a program tailored to anxiety in the first instance.

The result that users with higher dysfunctional thinking scores experienced less depression improvement was also unexpected. The CBT content within the *MoodGYM* program is concerned with identifying and challenging dysfunctional thoughts. Hence, it might be assumed that users with high levels of dysfunctional thinking would benefit most from the intervention. However, this was not the case. In order to check whether the finding was consistent across all three factors in the Warpy Thoughts scale, the multivariate regression model was rerun, replacing the Warpy Thoughts Score with scores for each of these three factors entered simultaneously (see Appendix B: Additional analyses, section B.3). All three variables were highly significant, with higher scores predicting less depression improvement, and overall there was no improvement in the model fit. Hence, we conclude that the findings were not confined to a specific dysfunctional thinking factor. The overall finding that higher dysfunctional thinking predicts less depression improvement does, however, reflect the results of a study which investigated predictors of depression improvement for participants who were randomised to receive *MoodGYM* with face-to-face guidance from a therapist [182]. In that study, the authors concluded that the intervention might have been more effective for those participants with lower levels of dysfunctional thinking due to an associated increased cognitive flexibility which allowed the participant to benefit from cognitive techniques [182].

Previous observational studies of *MoodGYM* users have focused on adherence and gender as predictors of depression improvement [130, 131, 132]. The current study extended these studies by examining all registration characteristics (including motivation and past experience of CBT and other demographic variables) as potential predictors of depression score change, in addition to clinical characteristics and adherence measures. Our finding that motivation to complete the program was a predictor of depression improvement is therefore novel, as are the observations that users who had either previously tried CBT and found it useful, or had not experienced CBT but wished to do so, were more likely to improve, and that those who had previously found CBT not useful were *less* likely to improve. These motivation and experience predictors suggest that the user's attitude to the program material may be an important determinant of the effectiveness of the intervention, perhaps because

these attitudes moderate how open the user is to engaging with and practising CBT concepts. In addition, past negative experience with CBT may indicate that the program is not an effective treatment type for that individual.

As has been reported in a previous study of outcomes for spontaneous *MoodGYM* users [132], female users experienced greater depression improvement than male users. In addition, we found that users who were more highly educated improved more, a finding that might be explained by superior written or verbal literacy and comprehension of the CBT concepts. Self-reported history of depression and use of anti-depressant medication both predicted less depression improvement. It could be that these users have chronic or treatment refractory symptoms or were experiencing recurrent depression symptoms which were difficult to treat.

Predictors based on the location of the user are not easily interpreted. Why did users from the United Kingdom experience less depression improvement, and why did those from rural areas demonstrate greater improvement? Perhaps another, unmeasured factor such as locus of control [156] mediated the relationship. Previous research suggests that individuals located in rural areas score higher on measures of self-reliance and stoicism [158], which raises the possibility that these individuals have a strong internal locus of control (greater self-belief in one's capacity for personal control over one's health). Given that use of the *MoodGYM* program requires users to be self-directed, and to practice concepts without human guidance, it is conceivable that users with a higher internal rather than external locus of control might be more likely to benefit from the intervention. Similarly, it might be hypothesised that users from the United Kingdom have a higher external locus of control (greater belief in the influence of external forces on one's health). More research is required on the locus of control of spontaneous community users in order to understand its impact on user outcomes and possible association with user location.

It is of interest that although being aged 19 years or younger predicted an increased number of modules completed, membership of this age group predicted *less* depression improvement, even before controlling for adherence. We cannot exclude the possibility that CBT in general is less effective for younger individuals, although evidence for the effect of age on response to CBT in face-to-face settings is mixed [183]. It is also possible that iCBT, or the *MoodGYM* program in particular, are less effective in younger compared to older users.

5.4.4 Limitations

There were several limitations with the current study.

First, the nature of the self-directed program and the collection of follow-up assessment data in real time (as part of the use of the intervention) meant that post-test score data for some participants was collected only a short period of time after their pre-test data. It seems unlikely that changes in depressive symptoms which are recorded within less than a day are valid or sustained, particularly given that the assessment asks, 'how you have been feeling *lately*'. It is also important to note that although the overall improvement in depression scores for users who completed two or more depression assessments was statistically significant, the mean improvement was only 0.58 points (95% CI: 0.57, 0.59) with a small effect size (Cohen's d = 0.28), which does not necessarily translate into clinical or meaningful change at an individual level. However, for a public health intervention a small individual effect can represent a major benefit in population terms.

In addition, given that users were required to repeat the depression assessment as they progressed through the program (potentially up to six times – at the start of each module and at the end of the program), we cannot exclude the possibility that the findings of improvement in depression are a 're-test artefact' given that reassessing depression symptoms may in itself result in reduced symptom scores [184]. Since users who complete more of the program complete more depression assessment surveys, it is possible that the greater improvement in symptoms for these users is confounded by this re-test artefact.

In addition, conclusions are limited by the lack of a control group or an indication of the likely improvement expected among participants who receive no treatment. Hence, user attributes which predict improvement in depressive symptoms in *MoodGYM* users might also be associated with spontaneous depression improvement in similar individuals who have not accessed the intervention. The analysis of change in depressive symptoms was also restricted to a 'completer' analysis of users who completed at least two depression assessments, which was only 38.4% of those users who commenced Module 1. As such, the analysis did not account for missing data from the majority of users who did not progress to the second depression assessment.

As noted in Chapter 1, it is also possible that users accessed other mental health services, on the recommendation of intervention feedback, and that depressive symptom improvement is attributable to those services. Without follow-up measures of help-seeking behaviours or service use it is not possible to know if users accessed other treatment options in the period between pre- and post-tests. However, the study presented in Chapter 4 found that the majority of this user sample

had previously accessed three or more evidence-based treatments for depression and approximately half indicated that they had not previously received CBT but wanted to try it. Hence, especially when combined with barriers to accessing face-to-face depression treatment, it is likely that many users did not actively pursue alternative services during their use of the *MoodGYM* program.

Finally, as discussed in the previous chapter, it is possible that we failed to collect data on all key predictor variables. Such variables might include demographic factors such as socio-economic, employment and relationship status and concurrent treatment, as well as personality factors such as locus of control. It was also a limitation that no data was collected to ascertain whether an individual user was receiving support from an external person, and the registration and clinical data were self-reported by anonymous users, with no means to confirm its accuracy.

5.5 Conclusion

There is encouraging evidence that *MoodGYM* may be effective when implemented outside of a research trial environment. Although the current study is observational and improvement in depression symptoms could be attributed to factors other than the intervention, the results show that there is a small but significant reduction in depression symptoms in those real-world users who complete more than one depression assessment as part of their use of the program. Greater improvement was most strongly predicted by completion of more of the intervention, which emphasises the importance of user engagement, a well-recognised challenge for the success of self-directed interventions [127]. While few spontaneous community users completed the whole *MoodGYM* program, a large number of users (approximately 50,000) completed two or more of modules over the 5-year period, which indicates a meaningful public health impact. However, it is likely that this impact could be amplified if user engagement with the intervention could be increased.

Motivation to complete the program at the time of registration, and not previously having experienced CBT but wishing to try it, were both significant predictors of adherence to the program, and were also independent predictors of depression improvement after controlling for adherence variables. Hence, further work is needed on ways to motivate and educate interested users about iCBT and what to expect from a program such as *MoodGYM.* This pre-intervention could potentially include motivational interviewing techniques or brief information delivered via video or interactive

HTML and the use of adaptive programs. In addition, research is required on additional psycho-social factors which may predict adherence or outcomes. Finally, research on real-world uptake and user outcomes for *different* iCBT interventions is needed, in order to better understand which predictors of outcome are specific to *MoodGYM* and which may be generalisable to iCBT interventions.

6. Variations of MoodGYM: A Partially Randomised Preference Trial

6.1 Introduction

The previous chapters have shown that there is evidence that unguided depression iCBT interventions are effective for those who are recruited to RCTs from community settings, and that there are small but significant reductions in depression symptoms among real-world spontaneous community users of the *MoodGYM* automated iCBT program, particularly for those users who were motivated to complete the program.

However, as noted both in the current results and in the extant literature, adherence among spontaneous iCBT users is low. This raises the question as to whether it is possible to construct a shortened but still effective variant of *MoodGYM* that due to its reduced demands is accessed in full by more unguided users. One approach to addressing this question might be to redevelop the program with similar content but in shortened form. An alternative approach is to test *which* content from the *MoodGYM* program is required to achieve a reduction in symptoms. The study reported in this chapter adopts the latter approach. It examines the effectiveness for spontaneous community users of six variants of the *MoodGYM* program each comprising a different combination of the pre-existing program modules. In particular, the study seeks to investigate which of the following components are necessary for depression symptom reduction - 'brief cognitive therapy', 'extended cognitive therapy', 'behavioural strategies', 'stress reduction' and 'problem-solving strategies'. In addition, the study compares the effect of content on adherence and user outcomes, depending on whether the user chose to select their preferred program.

6.1.1 The effect of treatment preference and trial design on adherence and outcomes

An RCT compares different experimental conditions through random assignment, and in some cases stratification of participants across key baseline characteristics such as age, gender or symptoms, with the aim of minimising extraneous differences between the groups. However, participant preferences for allocation to specific conditions also potentially affect the internal validity and hence the results of RCTs. Such preferences may affect an individual's compliance with or acceptance of their assigned condition (depending on the interaction between their awareness and attitudes to the treatment options and their allocated condition), as well as their engagement with or attrition from the trial [185]. Preferences may also be a barrier to potential participants agreeing to be randomised in a trial, biasing the sample and limiting the generalisation of the results [186].

One way of addressing the potential biases associated with patient preference in RCTs is to employ a Partially Randomised Preference Trial (PRPT) design [185, 187]. The concept of a PRPT (also known as a Comprehensive Cohort Design) was first introduced by Brewin and Bradley in 1989. Within a PRPT design, participants who prefer a particular treatment are permitted to choose this treatment, and those without a preference are randomised to a treatment condition [188]. This design maximises the number of eligible participants who are included in the study and allows for the control of preference effects [189].

The literature on treatment and management of mental health conditions suggests that clients often have strong preferences about the treatments that they receive. Two meta-analyses involving face-to-face services suggest that this preference impacts on an individual's motivation to engage with a prescribed treatment and on the outcomes of the treatment [190].

In a 2009 meta-analysis of 26 studies comparing clinical outcomes for study participants who received their preferred mental health treatment compared with participants who did not, Swift and Callahan found a small but significantly better outcome for those assigned to their preferred treatment [191]. In addition, participants matched to their preferred treatment were less likely to drop out. The meta-analysis included nine RCTs in which participants were asked about their preferences and then randomised to either match or not match that preference and eight RCTs where preference was assessed but not taken into consideration for random assignment to conditions. Additionally, the meta-analysis included six trials where patients with strong preferences were provided with their choice of treatments using a PRPT design. The effect of client preference on outcomes was significantly lower for the PRPTs than for the other included trials, a finding that might have been predicted given that in the PRPT designs no participant received a treatment if they had a strong preference for an alternate treatment [191].

A more recent meta-analysis studied the effect of client preference on clinical outcomes and treatment completion for RCTs of behavioural health interventions for mental disorders, as well as trials of interventions for other medical conditions; the analysis excluded PRPTs [192]. This analysis also found small positive aggregate effects of preference on outcomes and treatment completion. These effects were found to be reliable across sub-groups, including study design (whether the participant was randomised to choose their preferred treatment or if their preference was simply assessed), the information provided (whether participants were educated about their treatment choices), and target disorder (mental or non-mental disorders).

There have also been a small number of studies of the effect of user preference on the effectiveness of e-health interventions. Two of these studies employed controlled designs. In a large participant preference study, Muñoz and colleagues compared outcomes for participants who were permitted to choose from all elements of an open-access smoking cessation intervention with the outcomes of those who had been randomised to receive particular components as part of previous RCTs [193]. They found higher smoking cessation rates for those participants provided with a choice of intervention [193]. However, in a study of the effectiveness of a web-based physical activity intervention, Vandelanotte and colleagues found that whether or not a participants' assigned condition was matched to their preference for intervention delivery by text, video or both text and video, had no effect on outcomes [194]. It is unclear if the disparity in findings for these two studies reflects differences between the targeted conditions, differences in the focus of preference (content vs presentation media) or some other factor.

In addition, two uncontrolled studies have examined the effectiveness of mental health Internet interventions which incorporated user preference. In the first, Andersson and colleagues conducted a pilot study of iCBT for anxiety disorders in which participants could choose which intervention modules they completed (but the first two psycho-education modules were compulsory, as was the last module about relapse prevention) [195]. The study found significant improvements in anxiety symptoms in users; however, there was no control condition comprising participants who were not permitted to choose modules. In the second study of treatments for depression, Johnasson et al. compared an Internet-delivered psychodynamic therapy program to an iCBT program, with participants choosing and receiving their preferred intervention and indicating the strength of their preference [196]. Both interventions resulted in comparable, significant improvements in depression (measured on the BDI-II). Preference strength was significantly associated with depression outcome at 7-month follow-up for the iCBT condition only. The association between preference strength and adherence to the intervention approached significance for the combined iCBT and psychodynamic group ($p < .07$). However, in this study, all participants chose and received their preferred treatment, and hence, there was no comparison between outcomes and adherence for participants who chose their preferred intervention compared to those who did not.

To date there have been no studies of the effect of the receipt of preferred treatment on the outcomes for online depression interventions. Given this, and the broader paucity of research on the role of preferences in the context of iCBT interventions, the current study seeks to investigate whether allowing users to choose the variant of an iCBT intervention for depression results in superior adherence and outcomes compared to users who are randomised to a condition. To the

best of our knowledge, there are no Internet intervention studies of the characteristics of individuals who choose their treatment condition compared to those who are willing to be randomised. Hence the current study also examines whether there are demographic or clinical predictors of choosing an iCBT condition rather than agreeing to randomisation.

6.1.2 The effectiveness of different MoodGYM components

As noted in Chapter 5, the adherence rates for spontaneous iCBT users are low, with few *MoodGYM* users completing the whole program. In order to investigate whether a shortened variant of the program is effective, one approach is to explore which program content is required for the reduction of depression symptoms.

To date, only one paper has examined this question [118]. This previously published study investigated the outcomes of randomising spontaneous *MoodGYM* users to different combinations of *MoodGYM* modules [118]. The study analysed data from a sample of trial participants (*n* = 2,794) who registered with the *MoodGYM* program between 13 January and 26 May 2005, and who agreed to be randomised to one of six conditions. The conditions were combinations of the program modules: Module 1 only ('Brief CBT'); Modules 1 and 5 ('Brief CBT plus problem-solving'); Modules 1, 4, and 5 ('Brief CBT plus stress management and problem-solving'; Modules 1, 2 and 5 ('Extended CBT plus problem-solving'); Modules 1, 2, 3 and 5 ('Extended CBT plus behavioural strategies and problem-solving'); and all modules ('Full program'). The primary outcome measure was the score on the Goldberg Depression scale throughout the participant's use of the program. A mixed model repeated-measures analysis of variance was undertaken, and the largest improvement was associated with the condition which included extended CBT plus problem-solving (i.e., included Modules 1, 2 and 5). Contrasts between conditions showed that there was a significantly larger reduction in depression symptoms for conditions which included extended CBT compared to those which did not, except for the full program which did not result in significantly different change compared to the condition comprising Module 1 only. The study authors concluded that extended CBT may be the most important aspect of the *MoodGYM* program.

6.1.3 Current study

The current study examined the effectiveness of different *MoodGYM* module combinations, and included the same data that was reported in the previous trial of *MoodGYM* module combinations [118] but extended that study in several ways.

First, the randomisation of participants was conducted in the context of a PPRT. Hence the current study considered participants who were randomised, but also data from participants who selected their condition, as per the full trial design. This design enabled the role of preference on user adherence and outcomes to be investigated.

Second, the current study considered the full sample of participants who were randomised. The published paper was restricted to an analysis of participants who registered up until 26 May 2005; however, trial recruitment continued until 31 December 2005. In total, 7,794 participants were randomised, and it is this complete sample that is the subject of the current study.

The candidate contributed to the planning, set-up and trial management for the previous study and was a co-author on the published paper [118][6]. The current study was undertaken by the candidate independently of the original investigation with the guidance of the candidate's supervisory panel.

Study aims

The current study has the following aims:

Aim 1 a): *To investigate whether demographic or clinical characteristics of users predict if a participant would opt to choose a condition rather than agree to randomisation;*

b): *To compare adherence for participants who chose their condition with those of participants who were randomised;* and

c): *To compare depression outcomes for participants who chose their condition with those of participants who were randomised.*

Aim 2: *To investigate which program variants yielded improvements in depression symptoms amongst participants who were randomised to a condition (RCT arm only).*

The investigation of the relative effectiveness of different program variants (Aim 2) was restricted to the data from randomised participants. Although PRPT design is an effective method for investigating and accounting for user treatment preference (as is examined in Aims 1 a, b & c), it is not considered a 'gold standard' design for effectiveness studies. As noted by Gemmell and colleagues, using PRPTs to evaluate effectiveness can lead to biased results because of many potentially confounding variables [197]. Although there is a potential sample bias associated with RCTs, and potential user preference effects (as discussed above), the randomisation process seeks to ensure that any

[6] Note that this co-authorship is under the candidate's maiden name (Brittliffe).

confounding variables (whether identified or not) are equally distributed across conditions, enabling differences to be attributed to the effect of the intervention. Accordingly, the RCT arm of the study informed the effectiveness investigation.

6.2 Method

The study received approval from the Australian National University Human Research Ethics Committee (Protocol Number 2004/297), and the randomisation arm of the trial was registered with the International Standard Randomised Controlled Trial Number Register (ISRCTN 03012627).

6.2.1 Participants

Inclusion criteria

Eligible participants were personal, spontaneous community users who registered an account on the Version 2 English-language release of the *MoodGYM* program between 14 January 2005 and 31 December 2005, and who consented to participate in the trial.

Exclusion criteria

Users who answered 'yes' to any of the following questions during registration were deemed to be a non-personal user and excluded from the trial and directed to the full *MoodGYM* program: "I am a psychiatrist/ psychologist/ therapist who treats people with depression or anxiety"; "I am a researcher reviewing depression or anxiety sites"; "I am studying anxiety and depression as part of a college or university course".

After registration, personal users were asked if they were recommended to use the program by a health professional, and if they answered yes, they were also excluded from the trial and directed to the full *MoodGYM* program.

Users were then asked if they were willing to assist in the evaluation of the current variants of the program and provided with general research trial information, including privacy and confidentiality procedures. Users consented to participate in the trial by clicking 'I agree (If you agree, this means that you will be helping to evaluate which of our variants of the program works best)'. Users who did not consent were excluded from the trial and directed to the full *MoodGYM* program.

6.2.2 Choice or randomisation

All eligible users who agreed to participate were asked if they would like to choose a variant of the program or be randomised, as per the wording in Figure 6.1.

There are six versions of the program we are evaluating:

- A version which provides the program but very briefly
- A version which provides the brief program but also problem-solving
- A version which provides the brief program but also includes stress management and problem-solving
- A version which provides the program but also problem-solving
- A version which covers everything except stress management
- A version which covers all of the program material in full

You can let us choose one of these programs for you at random or you can choose one of these programs now.

Choose for me / I want to choose

Figure 6.1. Text displayed to eligible trial participants.

Individuals who were willing to be randomised to the trial conditions were assigned one of the trial conditions using an automated randomisation algorithm, stratified by age and gender to ensure the even distribution of users across the six conditions in the following four demographic groups: females under 30 years; females 30 years and over; males under 30 years; and males 30 years and over. Users who indicated that they would prefer to choose a condition were able to select their preferred condition.

Once a user was either randomised to a condition or chose a condition, they were directed to the initial depression, anxiety and dysfunctional thinking assessments followed by the first module. Later modules which were included in their condition became available as they progressed through the intervention. The trial protocol did not introduce any emails or other contact for guidance or follow-up of users.

6.2.3 *Conditions*

The intervention conditions were variations of combinations of the existing *MoodGYM* program modules described in Chapter 1.

The six conditions were:

- Condition 1: 'Brief CBT' – Module 1 only
- Condition 2: 'Brief CBT plus problem-solving' – Modules 1 and 5.
- Condition 3: 'Brief CBT plus stress management and problem-solving' – Modules 1, 4 and 5.
- Condition 4: 'Extended CBT plus problem-solving' – Modules 1, 2 and 5.
- Condition 5: 'Extended CBT plus behavioural strategies and problem-solving' – Modules 1, 2, 3, and 5
- Condition 6: 'Full program' – All five modules.

Module 1 provided an introduction to CBT concepts, demonstrating the way in which mood is influenced by thinking; Module 2 extended this by describing types of dysfunctional thinking and methods of challenging them; Module 3 taught behavioural methods to overcome dysfunctional thinking, including activity scheduling and sections on self-esteem and assertiveness; Module 4 assessed life-event stress and provided information about relaxation techniques and stress management; and Module 5 focused on relationship break-ups and simple problem-solving approaches.

6.2.4 *Measures*

Registration data

Self-report data were collected from the registration form for each user. This registration information included gender, age group, country of registration, location in a rural/remote region, education level and previous history of depression. Unlike the data collected in the studies reported in Chapters 4 and 5, motivation level and previous experience of CBT were not included in the registration questions. There were also some differences in the response categories for age group and education level compared to the previously reported data, the current data having fewer categories. In addition, prior to analysis, the country of registration responses were re-coded into continents (Oceania, North America, Europe, Asia, Africa, South America).

Depression and anxiety symptom levels

As described in the previous chapters, users completed the 9-point Goldberg depression and anxiety symptom surveys before commencing Module 1 [136]. This survey was repeated at the beginning of any subsequent modules delivered as part of the intervention condition, and then at the end of the program.

Warpy Thoughts Quiz

In order to proceed to Module 1, users were required to complete the Warpy Thoughts Quiz, which measures dysfunctional thinking. As described in Chapter 5, psychometric evaluation has demonstrated that 20 of the 42 items in this scale contribute to three factors (relationships, entitlement and achievements) and a higher order factor from which a Warpy Thoughts score can be obtained [137]. The resulting score can range from 0 to 80, with higher scores reflecting higher levels of dysfunctional thinking. Given its demonstrated factor validity, the score for the current study was based on this 20-item version of measure [137].

Adherence measures

The primary measure of adherence was whether the user completed all the modules in their chosen or assigned condition. The number of modules completed by the user was also examined.

In addition, given that the conditions varied substantially in length (from one module to five modules), in order to examine adherence in a more standardised way, engagement with the program was also measured based on whether the user had completed at least two depression assessments (Yes: ≥ 2 assessments, No: < 2 assessments). For users in Condition 1 (Module 1 only) this comprised the initial depression assessment as well as the final depression scale at the end of the program (which became available after they completed Module 1). For users in all other conditions, this comprised the initial assessment as well as the depression scale at the beginning of the second module in their condition.

6.2.5 Analyses

All statistical analyses were conducted using SPSS Version 25.0 [140].

Aim 1: Predictors, adherence and depression outcomes for choosing versus accepting randomisation

Aim 1a: Predictors for choosing versus agreeing to be randomised

The number and percentage of participants who chose or were randomised were calculated. Univariate analyses were conducted to investigate the association between user 'choice status'

(whether a participant chose a condition or accepted randomisation) and their registration characteristics, symptom levels and level of dysfunctional thinking. Chi-square statistics were calculated for categorical variables, and for variables with more than two categories pairwise post hoc tests were undertaken using a Bonferroni correction to account for multiple comparisons. Independent samples *t*-tests were performed for continuous measures, and Levene's tests for equality of variances were undertaken. Where the assumption of homogeneity of variance was violated, Welch's *t*-test was reported.

A logistic regression was then conducted in order to test whether 'choosing' could be predicted by these registration and initial assessment variables. The dependent variable was 'Chose condition' (Yes, No), coded 'Yes' if the user chose their program and coded 'No' if they agreed to be randomised.

The independent variables included in the model were:

- Gender: female [0, 1];
- Age group: Dummy coded with ≥ 45 years as the reference group, ≤ 19 years [0, 1], 20-29 years [0, 1], 30-44 years [0, 1];
- Continent: Dummy coded with Oceania as the reference group – Europe [0, 1], North America [0, 1], Other [0,1] – the latter created by collapsing Asia, Africa and South America into a single category.
- Location: rural [0, 1];
- Education level: Dummy coded with Higher degree as the reference group, Some/no secondary [0, 1], Secondary only [0, 1], Undergraduate [0, 1], Bachelor's degree [0, 1].
- Previous history of depression: Depression history [0, 1];
- Baseline depression: Goldberg depression score [range 0 to 9];
- Baseline anxiety: Goldberg anxiety score [range 0 to 9];
- Baseline dysfunctional thinking: Warpy Thoughts Score [range 0 to 80].

The independent variables were entered into the model simultaneously. To check for multicollinearity, each independent variable was regressed against the other independent variables, and the Variance Inflation Factors (VIFs) for each of the variables in the regressions inspected [174].

Aim 1b: Adherence outcomes for choosing versus accepting randomisation

Of those participants who completed the initial assessments and commenced Module 1, the percentage of participants who completed their condition was calculated, based on whether they chose their condition or were randomised. Chi-square statistics were undertaken to determine whether there was a significant difference in the proportion of users who completed their intervention based on their choice status. These proportions were also calculated for each of the six conditions, and separate Chi-square statistics calculated to check for significant differences within each condition.

A binary logistic regression was then conducted to investigate if completion predicted user choice status after controlling for registration, initial assessment variables and condition. The independent variables were entered simultaneously and comprised those from the logistic regression in the previous section (see Aim 1a), as well as choice status [0, 1].

The number of modules completed by participants in each condition was also examined, with the proportions of users who chose and who were randomised compared using Chi-square tests.

In order to account for the varying lengths of the different conditions, the percentage of participants in each condition who completed at least two depression tests were then computed, according to whether or not participants chose their condition. Chi-square statistics were undertaken to test for significant differences in the proportions who fell in the chose category compared to the randomised category, both overall and within each condition.

An additional logistic regression was conducted with completion of at least two depression tests as the dependent variable. In order to test if choice was a significant predictor after controlling for other user characteristics, and then subsequently controlling for the condition which was either chosen or assigned to the user, a hierarchical logistic regression was undertaken. Independent variables were added in two blocks, with choice status, registration variables and initial assessment scores in the first step (Model 1) and condition added in the second step (Model 2). Model fit tests were performed, and multicollinearity was checked by regressing each independent variable against the other independent variables and inspecting the VIFs for the variables in each regression.

Aim 1c: Depression outcomes for choosing versus accepting randomisation

In order to examine the changes in depression scores for users based on choice status, the mean baseline and post-test depression scores were calculated for each group, along with the change in depression scores, for those participants who completed at least two depression tests. The post-test

depression score for a user was defined as the repeated measure recorded at the most advanced point in the program. The change in depression symptom levels was calculated as the baseline score minus the post-test score so that a positive change reflected an improvement in depression symptoms.

An ANCOVA analysis was undertaken to investigate the effect of choice status on depression change score after controlling for initial depression score for this 'completers' sample. Hierarchical regression analysis was then performed to determine whether choice status was a significant predictor of depression change scores for these users, after controlling for baseline assessment variables, registration variables and the condition which the user either chose or was randomised to receive. The assessment and registration variables were added in the first block, followed by choice status, and condition was added as the final block.

Aim 2: Depression outcomes for different program variants

Analyses of depression outcomes for the different combinations of *MoodGYM* modules were undertaken within the RCT arm of the study and were conducted on a modified intent-to-treat basis, with all randomised participants included in the analyses if they were deemed to have received the intervention. Participants were classified as having received the intervention if they completed the initial depression assessment (administered before the commencement of the first module), and so for whom, at least one set of outcome data was collected. Users who were classified as having received the intervention were included in the analysis regardless of how much they engaged with the program.

The primary outcome measure was the Goldberg Depression scale score, which was administered potentially six times (before each module and at the end of the program). Given that the different conditions in the trial design omitted different modules, not all assessments were administered to all participants (planned missingness). Participants were required to complete their assigned modules and associated scales in linear order. Hence, if a participant failed to complete an assigned assessment, all further data were missing for that participant (i.e., the participant dropped out of the study permanently which means that missingness followed a monotone pattern [198]).

The demographic and clinical characteristics of participants in the six conditions were examined, and Chi-square and ANOVA tests performed to check for any significant differences between the groups. The number of modules and the number of assessments completed by participants in each group were also inspected.

A linear mixed effects model for repeated measures was then undertaken. This model enables the use of data from all participants who received the intervention (not only those who completed the post-test assessment) and assumes that any missing data is 'missing at random' (MAR) – that is, that missing data may be related to the participant or the intervention characteristics, but not to the value of the data which is missing. This ITT approach is the preferred approach to analysing RCT data in intervention effectiveness research [122].

The mixed model included the Goldberg Depression score as the repeated outcome variable, and condition and measurement occasion (time) as the factors. The model included fixed effects for condition, time and the interaction between condition and time. Different residual structures for the model were tested, and an unstructured covariance structure was chosen based on model fit (determined through inspection of differences in -2 Restricted Log Likelihood, Akaike's Information Criterion and Schwarz's Bayesian Criterion). The estimated marginal means produced by the model for Goldberg Depression were then examined.

Between-groups effect sizes were calculated for each of the conditions compared to Condition 1 (Module 1 only), based on the difference in the estimated end-point depression means for each condition and the pooled standard deviation of depression scores at baseline. Contrasts between conditions for changes from baseline to post-test were also calculated. These contrasts were conducted for each of the conditions compared to Condition 1, and between the combined conditions that included each of other modules and those that did not.

6.3 Results

During the trial period, 41,283 spontaneous community users registered with the *MoodGYM* program. Of these users, 3,190 were ineligible for participation in the trial because they were non-personal users who answered 'yes' to being a health professional, site reviewer or student studying CBT (*n* = 3,180) or were test users (*n* = 17). An additional 6,628 users were not offered participation in the trial because they indicated that they were referred to the program by a health professional, and so were directed to the full program.

A total of 31,458 users were invited to be included in the study, of whom 23,988 consented to participate. Of these trial participants, approximately two thirds (*n* = 16,194, 67.5%) opted to choose their own program, whereas 7,794 (32.5%) agreed to be randomly assigned to a condition. This participant flow is described in Figure 6.2.

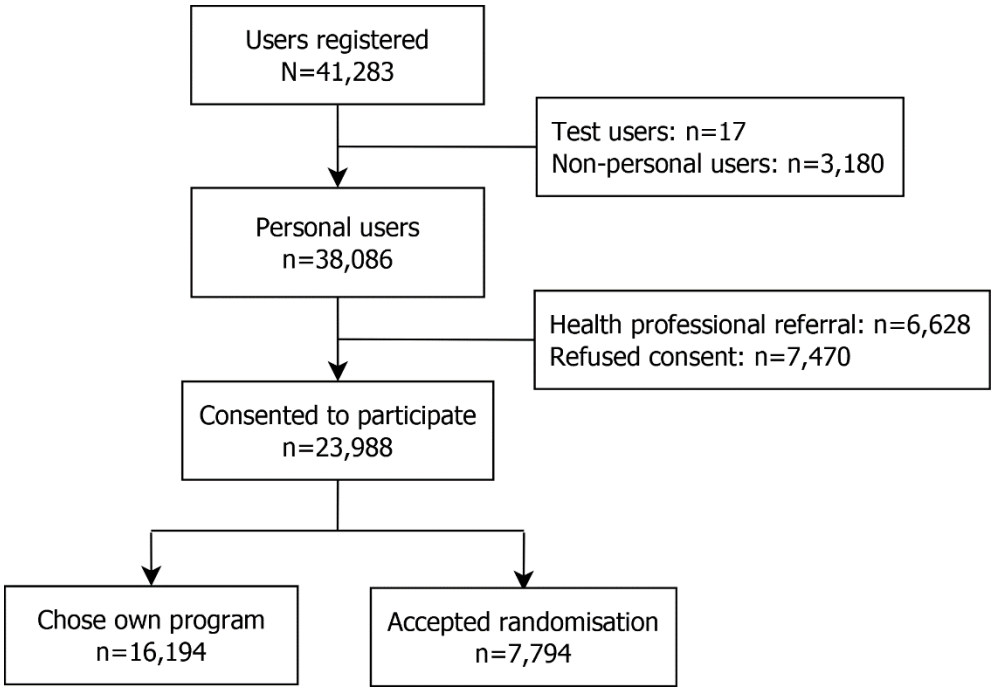

Figure 6.2. Number of users who were included and excluded from the trial (N=41,283 registered users).

Of those users who chose their own program, the majority (70.4%) chose Condition 6 which comprised the full program. The next popular choice was Condition 3 (18.4%) which included brief CBT, stress management and problem-solving. The numbers of participants who chose each condition are shown in Figure 6.3, along with the numbers who were randomised to each condition. The randomisation procedure (which included stratification across four demographic groups) resulted in roughly the same numbers of participants assigned to each of the six conditions.

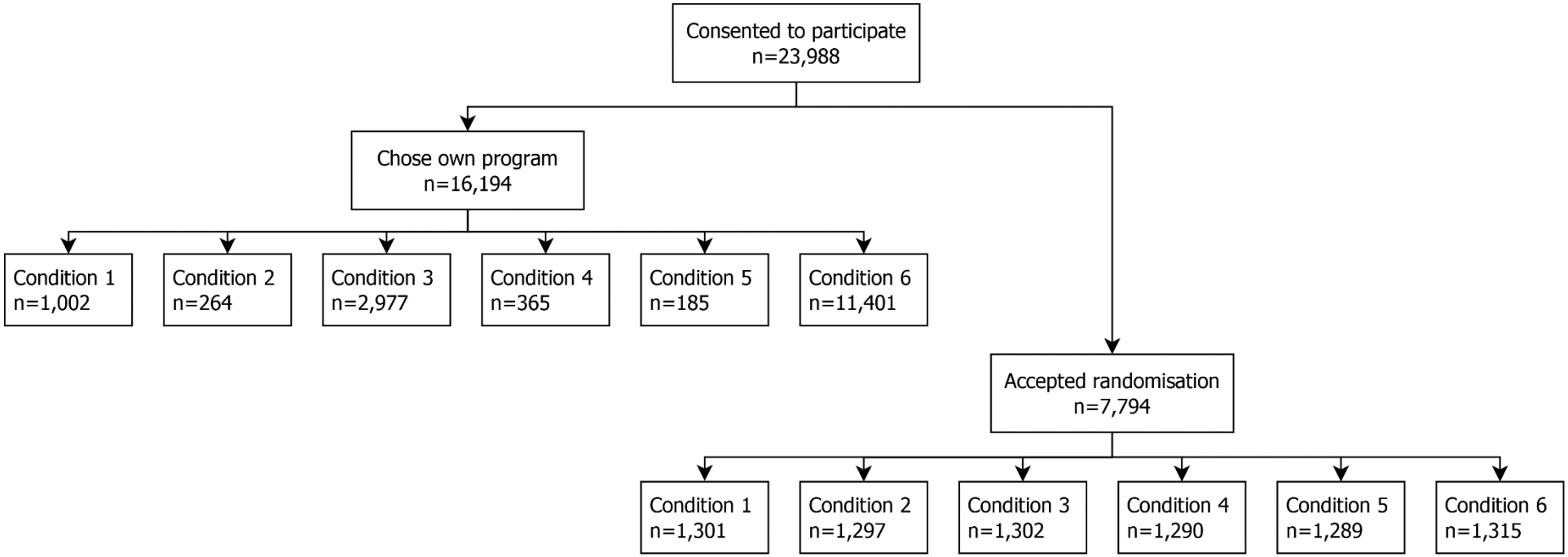

Figure 6.3. Number of trial participants who chose or were assigned to each condition (n=23,988).

6.3.1 *Aim 1: Predictors, adherence and depression outcomes for choosing versus accepting randomisation*

Aim 1a: Predictors of choosing versus accepting randomisation

The numbers of users who chose their program or who accepted randomisation are described in Table 6.1. The results of univariate analyses showed significant differences in the proportion of users who chose their program compared to those randomised for several registration variables. In particular, there were a higher proportion of users who chose their condition who: had a higher level of education; were older; were located in North America (compared to being in Oceania or Europe); reported previous use of anti-depressant medication; or had a self-reported history of depression. Gender and rural location were not associated with choice status.

A comparison of the continuous initial assessment variables for those who chose and those who were randomised demonstrated homogeneity of variances for depression and anxiety scores as assessed by Levene's test for equality of variances (p = .06, p = .64). However, Levene's test showed that the assumption of homogeneity of variances was violated for baseline dysfunctional thinking scores (p = .002), and hence Welch's t-test was used for this variable. Mean anxiety scores were 0.09 higher for users who chose their condition (95% CI: 0.02, 0.16). There was no significant difference in baseline scores between groups for depression symptoms or dysfunctional thinking.

Table 6.1. Choice status, by registration characteristics and mean baseline scores (n=23,988).

	Chose condition *n* (column %)	**Randomised *n* (column %)**	**Chi-square**	**Cramer's V**
Total	16,194 (100%)	7,794 (100%)		
Gender			1.78	0.009
Female	10,841 (66.1%)	5,150 (66.9%)		
Male	5,353 (33.1%)	2,644 (33.9%)		
Age group			**266.77*****	0.105
≤ 19 years[a]	1,408 (8.7%)	1,160 (14.9%)		
20 – 29 years[b]	4,561 (28.2%)	2,394 (30.7%)		
30 – 44 years[c]	6,131 (37.9%)	2,532 (32.5%)		
≥ 45 years[c]	4,094 (25.3%)	1,708 (21.9%)		
Continent			**140.08*****	0.076
Oceania[a]	8,606 (53.1%)	4,231 (54.3%)		
Europe[a]	2,609 (16.1%)	1,327 (17.0%)		
North America[b]	3,711 (22.9%)	1,368 (17.6%)		
Other[c]	1268 (7.8%)	868 (11.1%)		
Location			1.23	0.007
Rural / remote	2,911 (18.0%)	1,447 (18.6%)		
Urban	13,283 (82.0%)	6,347 (81.4%)		
Education level			**294.25*****	0.111
Some secondary[a]	2460 (15.2%)	1,774 (22.8%)		
Secondary only[b]	3,922 (24.2%)	2,012 (25.8%)		
Undergraduate[b]	2,487 (15.4%)	1,234 (15.8%)		
Bachelor's degree[c]	4,175 (25.8%)	1,614 (20.7%)		
Higher degree[c]	3,150 (19.5%)	1,160 (14.9%)		
Depression history			**9.58****	0.020
Previous depression	13,606 (84.0%)	6,425 (82.4%)		
No previous depression	2,588 (16.0%)	1,369 (17.6%)		
Anti-depressant use			**68.44*****	0.053
Previously used	8,139 (50.3%)	3,473 (44.6%)		
Not used	8,055 (49.7%)	4,321 (55.4%)		

Table 6.1. Choice status, by registration characteristics and mean baseline scores (n=23,988) (continued).

	Chose condition ***M (SD)***	**Randomised** ***M (SD)***	***t* (*df*)**
Depression (baseline) ***n* = 20,022**	5.95 (2.09) *n* = 13,719	6.02 (2.06) *n* = 6,303	1.66 (20,020)
Anxiety (baseline) ***n* = 19,631**	6.00 (2.21) *n* = 13,486	5.91 (2.21) *n* = 6,145	**2.69**** (19,629)
Warpy Thoughts (baseline) ***n* = 19,316**	43.71 (12.46) *n* = 13,273	43.84 (12.89) *n* = 6,043	0.65 (11,339.30)

*$p < .05$, **$p < .01$, ***$p < .001$.
For variables with more than 2 categories, each superscript letter denotes a subset of variable categories whose column proportions do not significantly differ from each other (after correction for multiple comparisons).

A logistic regression analysis was conducted to investigate whether any of the registration or initial assessment variables could predict whether a user chose their condition or agreed to randomisation. This model included data from the 19,316 participants who completed all of the initial assessments. All variables were entered simultaneously, and multicollinearity in the model was not problematic, with VIFs of less than 2 for all variables in separate regression analyses for each independent variable regressed against the other independent variables.

The model was statistically significant (χ^2 (17) = 424.71, $p < .001$); however, the Nagelkerke pseudo R^2 was 0.031 indicating that the independent variables only accounted for 3.1% of the variance (see Table 6.2). The Hosmer and Lemeshow test statistic was significant ($p < .001$) indicating that there was a significant difference between the observed values and those predicted by the model. This suggests a poor model fit, although this statistic is less reliable as a measure of goodness of fit in large samples such as the current dataset (*n* = 19,316) [142]. Overall the model had a moderate prediction success rate of 68.7% correctly categorised.

Table 6.2. Logistic regression model test statistics for choosing versus accepting randomisation.

	Chi-square (*df*, *p*)	**-2 Log likelihood**	**Cox & Snell R^2**	**Nagelkerke R^2**
Model (gender, age, continent, education, previous depression, initial depression, initial anxiety, initial warpy thoughts)	424.71 (17, <.001)	23,579.73	.022	.031

The odds ratios for each of the independent variables in the model are presented in Table 6.3, along with 95% confidence intervals, Wald statistics and the significance of the predictive contribution of each variable.

Table 6.3. Logistic regression results for predicting choosing condition versus accepting randomisation (n=19,316).

	Odds Ratio	95% CI	*P*	Wald
Gender				
Female	1.02	[0.95, 1.09]	.594	0.285
Age group ** reference group: ≥ 45 years*				
≤ 19 years	**0.68*****	[0.60, 0.77]	<.001	38.35
20 – 29 years	**0.80*****	[0.73, 0.87]	<.001	24.23
30 – 44 years	1.00	[0.92, 1.09]	.996	<.001
Continent* *reference group: Oceania*				
Europe	0.93	[0.85, 1.01]	.108	2.58
North America	**1.23****	[1.13, 1.33]	<.001	23.51
Other	**0.74*****	[0.66, 0.84]	<.001	23.69
Location				
Rural/remote	0.95	[0.87, 1.03]	.176	1.83
Education level ** reference group: Higher degree*				
Some secondary	**0.61*****	[0.54, 0.68]	<.001	72.62
Secondary only	**0.71*****	[0.64, 0.78]	<.001	45.76
Undergraduate	**0.78*****	[0.69, 0.87]	<.001	19.36
Bachelor's degree	0.97	[0.87, 1.07]	.494	0.468
Depression history				
Previous depression	**0.84****	[0.76, 0.93]	.001	10.76
Anti-depressant use				
Previously used	**1.16*****	[1.08, 1.24]	<.001	17.31
Depression (baseline)	**0.98****	[0.96, 0.99]	.009	6.82
Anxiety (baseline)	**1.04*****	[1.02, 1.05]	<.001	17.73
Warpy Thoughts (baseline)	1.00	[1.00, 1.00]	.723	0.126

**p* < .05, ***p* < .01, ****p* < .001

After controlling for other registration and initial assessment variables, education level was the most important predictor of choice versus randomisation. Users who had not completed secondary school were 39% more likely to accept randomisation than those with a higher degree, those who had completed secondary school only were 29% more likely, and current undergraduate students were 22% more likely. There were also much higher odds of younger users accepting randomisation

compared to older users, with users aged 19 years or under 32% more likely than those aged 45 years or older, and users aged 20-29 years 20% more likely. Users who were located in North America had 23% higher odds of choosing their program compared to those in Oceania, whereas there was no difference between those in Oceania and Europe. Users in other continents were 26% more likely to accept randomisation than those in Oceania.

Having used anti-depressants was a significant predictor of choosing (16% more likely); however, users with a self-reported history of depression were 16% more likely to accept randomisation. Higher initial depression scores were associated with being less likely to choose a program (2.4% more likely to be randomised with each unit increase in depression score), and higher anxiety scores were associated with being more likely to choose a program (3.6% more likely to choose with each unit increase in anxiety score).

Gender, rural location and initial dysfunctional thinking score were not significant predictors in the model.

Aim 1b: Adherence outcomes for choosing versus accepting randomisation

Completion of condition

Of the 19,316 participants who completed the initial assessments and commenced Module 1, 14.7% (*n* = 888) of the 6,043 participants who were randomised completed their assigned condition compared to 6.1% (*n* = 813) of the 13,273 participants who chose their condition. This difference was statistically significant ($\chi^2 = 379.72$, $p < .001$); however, this difference was strongly influenced by the large proportion of 'choose' participants who selected the full five module variant of the program. For the separate conditions, the difference in the proportion of participants who completed their program was not significantly different for any of the conditions except Condition 1 (Module 1 only). For Condition 1, a larger percentage of users who were randomised completed their program (41.3%) compared to those who chose this variant (36.5%; $\chi^2 = 3.23$, $p = .047$). The percentages of choose and randomised participants who completed each condition are displayed in Table 6.4.

Table 6.4. Numbers of participants in each condition who commenced Module 1 and completed their program, by choice status (n=19,316).

	Chose condition *n* (column %)	**Randomised *n* (column %)**	**Chi-square**	**Phi**
Condition 1			**3.23***	-0.05
Did not complete	454 (63.5%)	578 (58.7%)		
Completed	261 (36.5%)	406 (41.3%)		
Condition 2			2.89	-0.05
Did not complete	161 (81.3%)	761 (75.7%)		
Completed	37 (18.7%)	244 (24.3%)		
Condition 3			1.73	-0.02
Did not complete	2,096 (91.3%)	913 (89.9%)		
Completed	200 (8.7%)	103 (10.1%)		
Condition 4			0.02	-0.004
Did not complete	271 (91.9%)	920 (91.6%)		
Completed	24 (8.1%)	84 (8.4%)		
Condition 5			0.56	0.02
Did not complete	149 (96.1%)	978 (97.2%)		
Completed	6 (3.9%)	28 (2.8%)		
Condition 6			1.75	0.01
Did not complete	9,329 (97.0%)	1,005 (97.8%)		
Completed	285 (3.0%)	23 (2.2%)		
Total			**379.72*****	-0.14
Did not complete	12,460 (93.9%)	5,155 (85.3%)		
Completed	813 (6.1%)	888 (14.7%)		

$^{*}p < .05$, $^{***}p < .001$

A binary logistic regression analysis was performed to investigate if choosing a condition was a predictor of completing that condition. Potential registration and initial assessment predictors were also included in the regression; however, after controlling for the received condition, the regression model was not significant ($\chi^2 = 3.89$, $p = .56$). Hence program completion could not be predicted by choice status after controlling for condition.

Engagement with received condition

The number of modules completed by participants in each condition (of those that commenced Module 1), grouped by choice status, are shown in Figures 6.4 to 6.9 below. Chi-square tests showed that the difference in the proportion of participants completing different numbers of modules was significant for Conditions 1, 5 and 6.

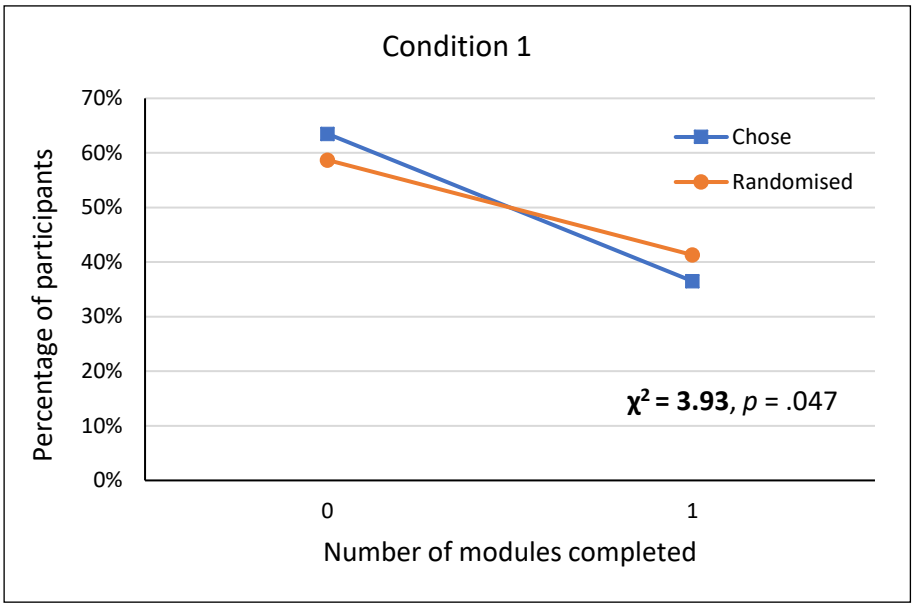

Figure 6.4. Number of modules completed by participants who received Condition 1, by choice status (chose: n=715; randomised: n=984).

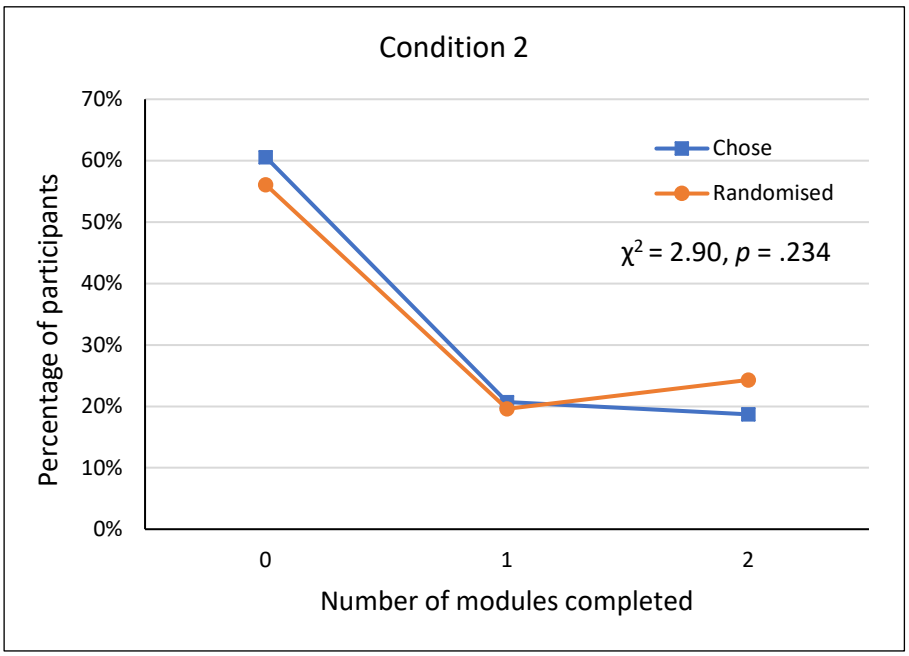

Figure 6.5. Number of modules completed by participants who received Condition 2, by choice status (chose: n=298; randomised: n=1,005).

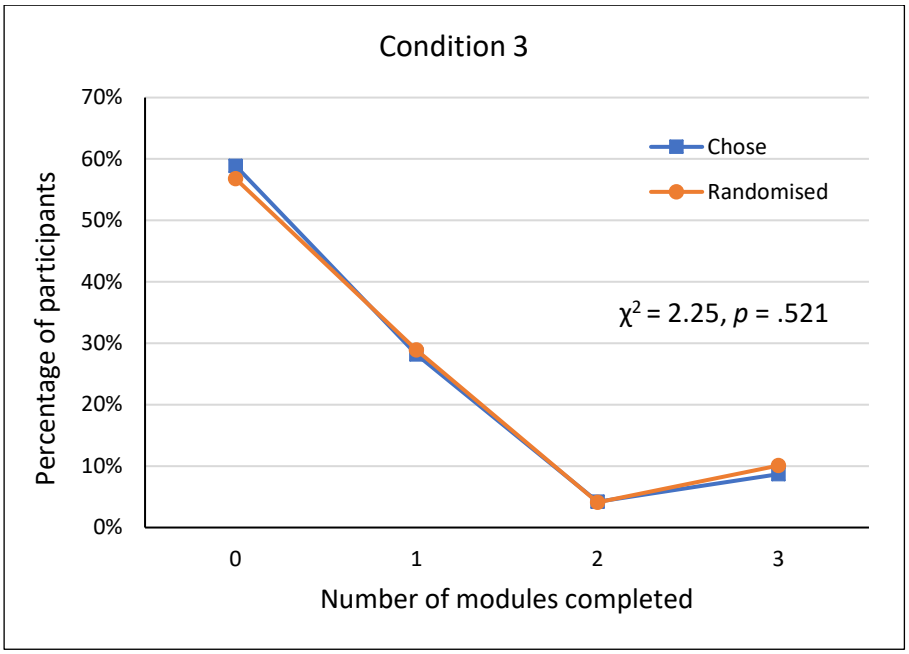

Figure 6.6. Number of modules completed by participants who received Condition 3, by choice status (chose: n=2,296; randomised: n=1,016).

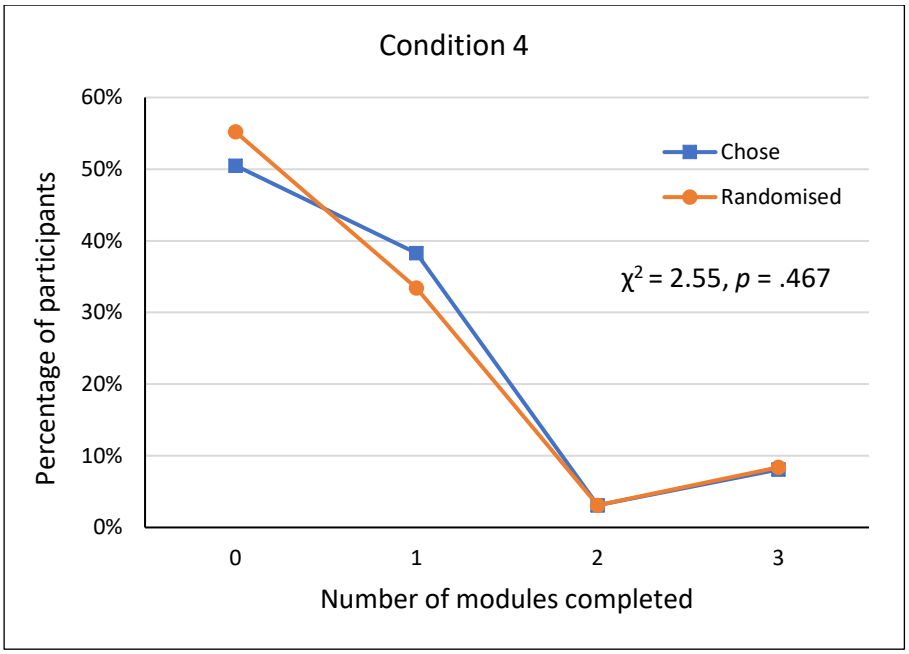

Figure 6.7. Number of modules completed by participants who received Condition 4, by choice status (chose: n=295; randomised: n=1,004).

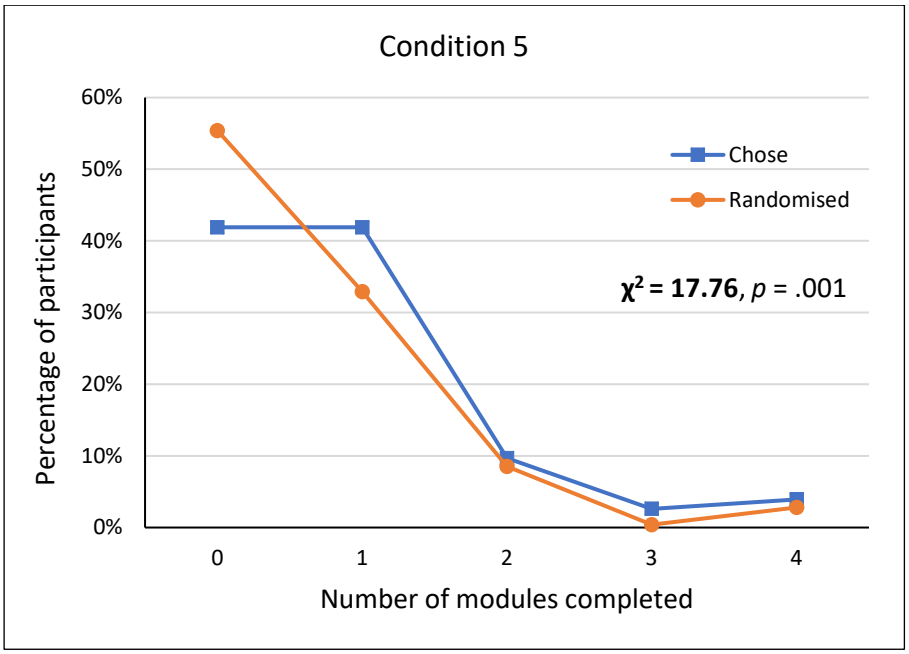

Figure 6.8. Number of modules completed by participants who received Condition 5, by choice status (chose: n=155; randomised: n=1,006).

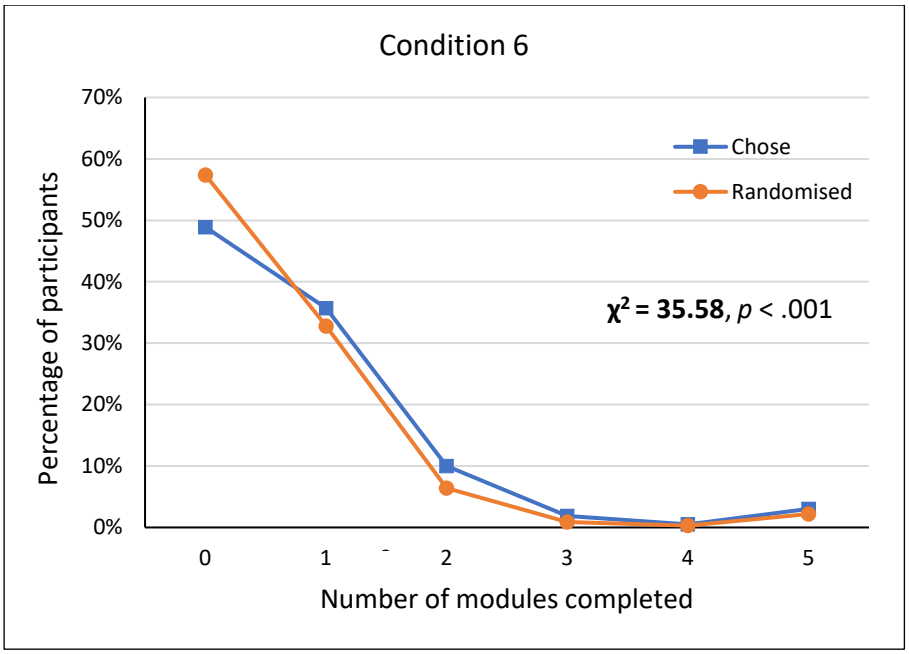

Figure 6.9. Number of modules completed by participants who received Condition 6, by choice status (chose: n=13,273; randomised: n=6,043).

In order to compare differences in engagement between those who chose and those who were randomised across the conditions with considerable variations in length, the numbers of participants who completed at least two depression assessments were examined.

Overall, there was a significant difference between choosers and those randomised, with a larger proportion of users who chose their condition (*n* = 4,465, 33.6%) completing at least two depression

assessments compared to those who were randomised (*n* = 1,746, 28.9%; $\chi^2 = 42.89$, $p < .001$). For specific conditions, there were significant differences in proportions who completed at least two depression assessments for Conditions 1, 2, 5 and 6 (see Table 6.5). For Conditions 1 and 2, a greater proportion of randomised users completed at least two depression assessments, and for Conditions 5 and 6, a greater proportion of users who chose their condition completed at least two depression assessments.

Table 6.5. Number of participants in each condition who completed at least two depression assessments, by choice status (n=19,316).

Number of assessments	Chose condition *n* (column %)	Randomised *n* (column %)	Chi-square	Phi
Condition 1			**11.53****	0.08
Less than two	574 (80.3%)	720 (73.2%)		
At least two	141 (19.7%)	264 (26.8%)		
Condition 2			**5.54***	0.07
Less than two	152 (76.8%)	687 (68.4%)		
At least two	46 (23.2%)	318 (31.6%)		
Condition 3			0.15	0.01
Less than two	1,651 (71.9%)	724 (71.3%)		
At least two	645 (28.1%)	292 (28.7%)		
Condition 4			0.58	-0.02
Less than two	203 (68.8%)	714 (71.1%)		
At least two	92 (31.2%)	290 (28.9%)		
Condition 5			**4.39***	-0.06
Less than two	95 (61.3%)	701 (69.7%)		
At least two	60 (38.7%)	305 (30.3%)		
Condition 6			**34.88*****	-0.06
Less than two	6,133 (63.8%)	751 (73.1%)		
At least two	3,481 (36.2%)	277 (26.9%)		
Total			**42.89*****	-0.05
Less than two	8,808 (66.4%)	4,297 (71.1%)		
At least two	4,465 (33.6%)	1,746 (28.9%)		

*$p < .05$, **$p < .01$, ***$p < .001$

A binary logistic regression analysis was undertaken to examine whether a participant's choice status was a significant predictor of completing at least two depression assessments, after controlling for other factors. Independent variables were entered in two blocks, with choice status, registration variables and initial assessment variables entered in Step 1 (Model 1) and the received condition

entered in Step 2 (Model 2). Multicollinearity within the model was checked through separate regression analyses for each independent variable (regressed against the other independent variables) and was not problematic with all VIFs less than 2.

Both models were statistically significant (Model 1: χ^2 (18) = 320.04, $p < .001$; Model 2: χ^2 (23) = 394.12, $p < .001$). The independent variables accounted for 2.3% of the variance in Model 1 and controlling for condition in Model 2 accounted for an additional 0.5% of the variance. The model test statistics are displayed in Table 6.6. The Hosmer and Lemeshow test statistics were non-significant for both models ($p = .62$; $p = .74$) indicating that there were no significant differences between the observed values and those predicted by the models, suggesting good model fit. Overall, both models had a prediction success rate of 67.8% correctly categorised.

Table 6.6. Logistic regression model test statistics for completion of two or more depression assessments.

	Chi-square (*df*, *p*)	**-2 Log likelihood**	**Cox & Snell R^2**	**Nagelkerke R^2**
Model 1 (choice status, gender, age, continent, location, education, depression history, anti-depressant use, baseline depression, baseline anxiety, baseline dysfunctional thinking)	330.04 (18, <.001)	23,942.01	.016	.023
Model 2 (Model 1 + condition)	394.12 (23, <.001)	23,867.94	.020	.028

The odds ratios for each of the independent variables in the models are presented in Table 6.7, along with 95% confidence intervals, Wald statistics and the significance of the predictive contribution of each variable.

In Model 1, after controlling for registration and initial assessment variables, choice status was an important and significant predictor of completing at least two depression tests, with those who chose their condition having 28% higher odds of completing two assessments. In addition, dysfunctional thinking score was an important predictor, with each unit in the Warpy Thoughts score resulting in 1.1% higher odds of completing two assessments. Other predictors were age group (with younger groups more likely to complete two assessments), continent (those from North America were less likely), depression history (those with self-reported depression history were more likely) and initial depression scale score (each unit increase on the initial Goldberg Depression scale resulted

in 3.4% greater odds of completing an additional assessment). Gender, education level, rural location, anti-depressant use and initial anxiety score were not significant predictors in the model.

Choice status was still a significant predictor of completing at least two depression tests, after controlling for the assigned or chosen condition in Model 2. In this model, users who chose their condition were 16% more likely to complete at least two depression tests. Users who were assigned or chose Condition 6 (all modules) were more likely to complete two assessments compared to those who received Conditions 1, 3 and 4. In addition, completion of two assessments was significantly predicted by higher initial dysfunctional thinking, younger age group, a history of depression, being located in Oceania or Europe, and higher baseline Goldberg Depression score.

Table 6.7. Logistic regression results for predicting completion of at least two depression tests.

	Model 1			**Model 2**		
	Odds ratio (95% CI)	***p***	**Wald**	**Odds ratio (95% CI)**	***p***	**Wald**
Choice status						
Chose condition	**1.28***** (1.20 – 1.37)	<.001	52.97	**1.16***** (1.07, 1.27)	<.001	11.72
Gender						
Female	1.04 (0.98 – 1.12)	.208	1.59	1.05 (0.98, 1.12)	.152	2.06
Age group ** reference group: ≥ 45 years*						
≤ 19 years	**1.41***** (1.24 – 1.60)	<.001	26.67	**1.38***** (1.22, 1.57)	<.001	25.01
20 - 29 years	**1.29***** (1.18 – 1.41)	<.001	30.64	**1.27***** (1.16, 1.39)	<.001	26.55
30 - 44 years	**1.15**** (1.06 – 1.25)	.001	10.45	**1.14**** (1.05, 1.24)	.003	9.00
Continent ** reference group: Oceania*						
Europe	**1.11*** (1.02 – 1.21)	.013	6.18	**1.11*** (1.02, 1.21)	.012	6.29
North America	**0.89**** (0.82 – 0.96)	.003	8.81	**0.88**** (0.81, 0.96)	.002	9.47
Other	1.02 (0.90 – 1.15)	.780	0.08	1.01 (0.89, 1.15)	.841	0.04
Location						
Rural/remote	1.03 (0.95 – 1.11)	.542	0.37	1.03 (0.95, 1.12)	.469	0.52

Table 6.7. Logistic regression results for predicting completion of at least two depression tests (continued).

	Model 1			Model 2		
	Odds ratio (95% CI)	***p***	**Wald**	**Odds ratio (95% CI)**	***p***	**Wald**
Education level ** reference group: Higher degree*						
Some secondary	0.96 (0.85 – 1.08)	.479	0.50	0.96 (0.85, 1.08)	.490	0.48
Secondary only	0.97 (0.88 – 1.07)	.530	0.39	0.97 (0.88, 1.07)	.592	0.29
Undergraduate	1.04 (0.93 – 1.16)	.486	0.49	1.04 (0.93, 1.16)	.456	0.56
Degree	1.03 (0.94 – 1.14)	.506	0.51	1.04 (0.94, 1.14)	.466	0.53
Depression history						
Previous depression	**1.23***** (1.11 – 1.37)	<.001	15.83	**1.22***** (1.10 – 1.35)	<.001	14.19
Anti-depressant use						
Previously used	0.95 (0.89 – 1.02)	.140	2.18	0.94 (0.88, 1.01)	.073	3.20
Depression (baseline)	**1.03***** (1.01, 1.05)	<.001	12.91	**1.03**** (1.01, 1.05)	.001	10.70
Anxiety (baseline)	1.01 (0.99, 1.02)	.409	0.68	1.01 (0.99, 1.02)	.395	0.72
Warpy Thoughts (baseline)	**1.011***** (1.008, 1.014)	<.001	62.67	**1.011***** (1.008, 1.013)	<.001	59.47
Condition ** reference group: Condition 6*						
Condition 1				**0.63***** (0.56, 0.72)	<.001	50.47
Condition 2				0.90 (0.78, 1.04)	.150	2.07
Condition 3				**0.77***** (0.71, 0.85)	<.001	32.42
Condition 4				**0.84*** (0.73, 0.97)	.016	5.77
Condition 5				0.94 (0.81, 1.09)	.397	0.72

**p* < .05, ***p* < .01, ****p* < .001

Aim 1c: Depression outcomes for choosing versus accepting randomisation

The mean baseline and post-test depression scores, and depression change scores are shown in Table 6.8 for the 6,211 participants who completed at least two depression tests (*n* = 4,465 chose condition; *n* = 1,746 randomised).

Table 6.8. Baseline, post-test and change in depression scores, by choice status (n = 6,211).

	Chose condition ***n* = 4,465** ***M (SD)***	**Randomised** ***n* = 1,746** ***M (SD)***
Baseline depression score	6.12 (2.01)	6.27 (1.88)
Post-test depression score	5.57 (2.38)	5.88 (2.20)
Change in depression score	0.56 (1.80)	0.38 (1.57)

An ANCOVA was run to investigate the effect of choosing versus being randomised on depression change score after controlling for initial depression score. After adjustment for initial depression score, the mean change in depression score for those who chose was 0.56 (95% CI: 0.51, 0.61) and 0.36 for those who were randomised (95% CI: 0.28, 0.44). This difference was statistically significant [$F(1, 6{,}208) = 17.20$, $p < .001$].

In order to further examine whether choice status was a significant predictor of depression score change after controlling for other variables, a hierarchical multivariate analysis was undertaken. Initial assessment scores (depression, anxiety and dysfunctional thinking) and registration variables were entered in the first block. Registration variables were retained in this block if their contribution to the model was statistically significant. This block included education level and previous history of depression, but gender, age group, continent and previous use of anti-depressants were all non-significant and removed from the model. Choice status was entered in the second block, and condition in the third block.

All models were statistically significant, and test statistics for each regression model are shown in Table 6.9. The final model accounted for 10.5% of the variance in the depression change score. Most of this was due to the initial assessment and registration variables, with choice status and condition each only accounting for an additional 0.2% of the variance. Multicollinearity was not problematic in the model, with all VIFs under 1.9. The Beta coefficients for the independent variables in each of the models are displayed in Table 6.10.

Table 6.9. Multivariate regression model test statistics for depression change score.

	R^2	R^2 change	*F* change	Sig *F* change
Model 1 (baseline depression, baseline anxiety, baseline dysfunctional thinking, education, depression history)	0.101	0.101	87.13	<.001
Model 2 (Model 1 + choice status)	0.103	0.002	11.52	.001
Model 2 (Model 2 + condition)	0.105	0.002	3.26	.006

Table 6.10. Hierarchical multivariate regression predicting depression change score (n=6,211).

	Step 1		Step 2		Step 3	
	β	*P*	β	*p*	β	*p*
Depression (baseline)	**0.349*****	<.001	**0.350*****	<.001	**0.348*****	<.001
Anxiety (baseline)	**-0.160*****	<.001	**-0.160*****	<.001	**-0.159*****	<.001
Warpy Thoughts (baseline)	**-0.118*****	<.001	**-0.117*****	<.001	**-0.118*****	<.001
Education level ** reference group: Higher degree*						
Some secondary	**-0.082*****	<.001	**-0.077*****	<.001	**-0.077*****	<.001
Secondary only	**-0.097*****	<.001	**-0.095*****	<.001	**-0.094*****	<.001
Undergraduate	**-0.048****	.003	**-0.046****	.004	**-0.046****	.003
Bachelor's degree	-0.010	.558	-0.010	.552	-0.010	.549
Depression history	**-0.061*****	<.001	**-0.060*****	<.001	**-0.060*****	<.001
Chose condition			**0.041****	.001	0.017	.288
Condition ** reference group: Condition 6*						
Condition 1					**-0.029***	.026
Condition 2					-0.025	.077
Condition 3					**-0.046*****	<.001
Condition 4					-0.006	.669
Condition 5					-0.021	.120

*p < .05, **p < .01, ***p < .001

Choosing a condition rather than agreeing to be randomised was a significant predictor of larger positive change in depression symptoms in Model 2 [Step 2: *B* = .159, β = .041, *t* = 3.39, *p* = .001]; however, its predictive contribution was not as important as baseline assessment scores, education level or history of depression. Higher baseline depression score, higher education level and history of

depression all predicted larger positive changes, whereas higher baseline anxiety and dysfunctional thinking scores predicted smaller positive changes.

However, choice status was no longer a significant predictor of depression change score in Model 3, which also controlled for the received condition. In this model, Conditions 1 and 3 significantly predicted less depression score change.

6.3.2 Aim 2: Relative effectiveness of different MoodGYM module combinations

In total, 7,794 participants agreed to be randomised and were allocated to one of the six conditions. Two-thirds of these participants were female (66.1%), 32.5% were aged between 30 and 44 years and more than a third (35.6%) had a Bachelor's or higher degree. The registration characteristics of these participants grouped by the condition to which they were randomised are shown in Table 6.11. Chi-square test statistics revealed no significant differences between the groups on any of these variables. The mean Goldberg Depression score for those participants who completed the initial depression assessment was 6.00 (*SD* = 2.06, *n* = 6,303), and an ANOVA analysis demonstrated that there was no significant difference in the mean baseline depression score between conditions.

Participants were classified as having received their intervention if they completed at least the baseline depression assessment. Overall, 6,303 (80.9%) of randomised participants submitted baseline depression data. The number of participants who received their intervention across conditions is shown in Figure 6.10, and there was an even distribution of participants who received their intervention across groups.

Table 6.11. Baseline characteristics and depression symptom scores for randomised participants, by condition (n=7,794).

	Condition 1 ***n*=1,301**	**Condition 2** ***n*=1,297**	**Condition 3** ***n*=1,302**	**Condition 4** ***n*=1,290**	**Condition 5** ***n*=1,289**	**Condition 6** ***n*=1,315**
	***n* (%)**	***n* (%)**	***n* (%)**	***n* (%)**	***n* (%)**	***n* (%)**
Female	862 (66.3%)	857 (66.1%)	858 (65.9%)	855 (66.3%)	846 (65.6%)	872 (66.3%)
Age group						
≤ 19 years	194 (14.9%)	184 (14.2%)	195 (15.0%)	173 (13.4%)	206 (16.0%)	208 (15.8%)
20-29 years	396 (30.4%)	411 (31.7%)	405 (31.1%)	408 (31.6%)	383 (29.7%)	391 (29.7%)
30-44 years	429 (33.0%)	415 (32.0%)	435 (33.4%)	411 (31.9%)	425 (33.0%)	417 (31.7%)
≥ 45 years	282 (21.7%)	287 (22.1%)	267 (20.5%)	298 (23.1%)	275 (21.3%)	299 (22.7%)
Rural location	264 (20.3%)	234 (18.0%)	218 (16.7%)	246 (19.1%)	233 (18.1%)	252 (19.2%)
Education level						
Some secondary	285 (21.9%)	287 (22.1%)	295 (22.7%)	287 (22.2%)	322 (25.0%)	298 (22.7%)
Secondary	342 (26.3%)	329 (25.4%)	333 (25.6%)	351 (27.2%)	314 (25.3%)	343 (27.2%)
Undergraduate	206 (15.8%)	221 (17.0%)	212 (16.3%)	212 (16.4%)	190 (15.3%)	193 (15.3%)
Bachelor's degree	274 (21.1%)	268 (20.7%)	273 (21.0%)	251 (19.5%)	272 (22.0%)	276 (21.9%)
Higher degree	194 (14.9%)	192 (14.8%)	189 (14.5%)	189 (14.7%)	191 (15.4%)	205 (16.3%)
History of depression	1,066 (81.9%)	1,080 (83.3%)	1,074 (82.6%)	1,073 (83.2%)	1,067 (82.8%)	1,065 (81.0%)
	M (SD) ***n***	***M (SD)*** ***n***	***M (SD)*** ***n***	***M (SD)*** ***n***	***M (SD)*** ***n***	***M (SD)*** ***n***
Depression (baseline)	6.1 (2.1)	6.0 (2.0)	6.0 (2.1)	6.0 (2.1)	6.0 (2.0)	6.0 (2.0)
	1037	**1043**	**1057**	**1035**	**1043**	**1088**
	5.9 (2.1)	5.8 (2.2)	6.1 (2.2)	5.8 (2.3)	5.9 (2.2)	5.7 (2.1)
Depression (end-point)	**264**	**318**	**292**	**290**	**305**	**277**

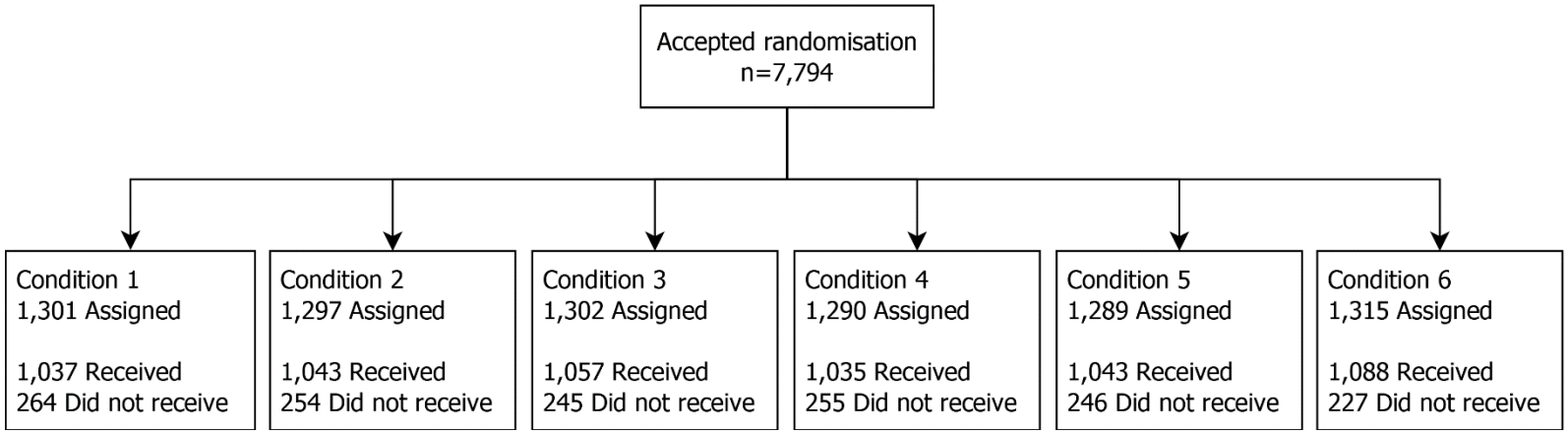

Figure 6.10. Numbers of randomised participants who received the intervention, by condition (n=7,794).

Of those users who received their intervention, most users (58.4%) did not complete a module, and 30.1% completed one module. The number of modules completed by participants in each of the conditions is displayed in Figure 6.11.

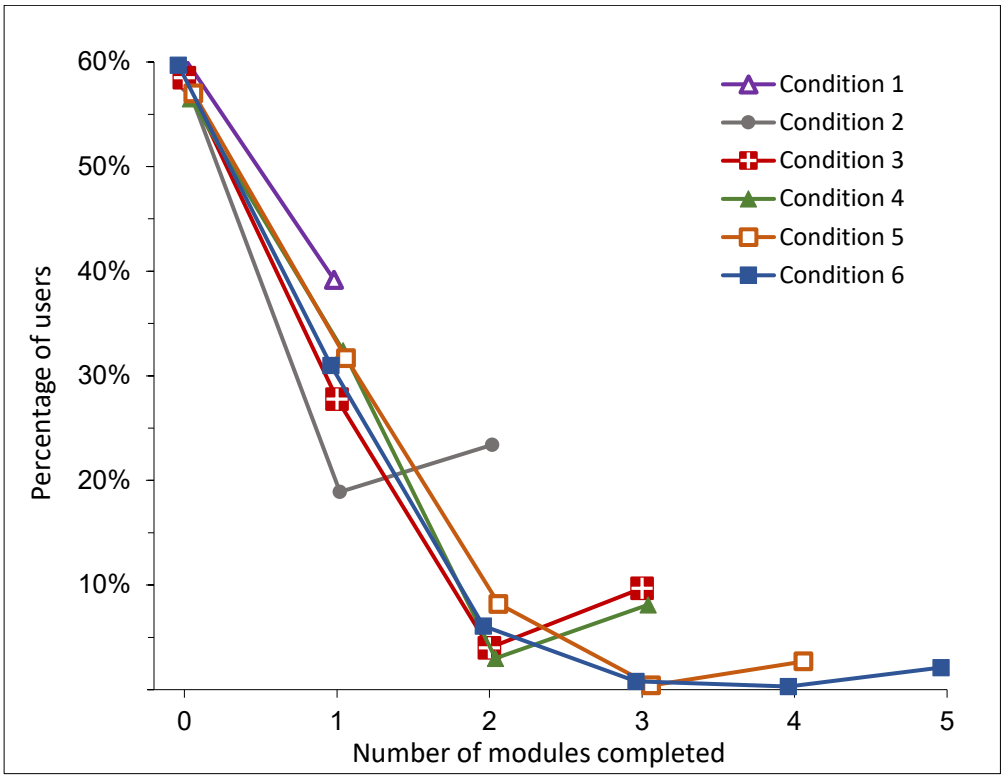

Figure 6.11. Number of modules completed by participants who received their condition (n=6,303).

Similarly, the majority of participants (72.3%) completed one depression scale only. The number and percentage of participants who completed the depression assessments which were available as part of their condition is shown in Table 6.12.

Table 6.12. Number and percentage of participants completing available depression assessments, by condition (n=6,303).

	Baseline ***n***	**Module 2** ***n* (%)**	**Module 3** ***n* (%)**	**Module 4** ***n* (%)**	**Module 5** ***n* (%)**	**Final** ***n* (%)**
Condition 1	1,037					264 (25.5%)
Condition 2	1,043				318 (30.5%)	190 (18.2%)
Condition 3	1,057			292 (27.6%)	125 (11.8%)	82 (7.8%)
Condition 4	1,035	290 (28.0%)			94 (9.1%)	71 (6.9%)
Condition 5	1,043	305 (29.2%)	80 (7.7%)		30 (2.9%)	25 (2.4%)
Condition 6	1,088	277 (25.5%)	73 (6.7%)	32 (2.9%)	24 (2.2%)	22 (2.0%)
Total	6,303	872 (13.8%)	153 (2.4%)	324 (5.1%)	591 (9.4%)	654 (10.4%)

A linear mixed effects model for repeated measures was fitted to the data, using an unstructured covariance structure which produced the best model fit statistics (-2 Restricted Log Likelihood = 36,069.76). The interaction of time and condition was statistically significant [$F(13, 627.44) = 2.34$, $p = .005$]. Figure 6.12 displays the estimated marginal means for each condition with error bars representing 95% confidence intervals. Visual inspection suggests that participants in Conditions 4 and 5 experienced the best depression outcomes, although differences were small.

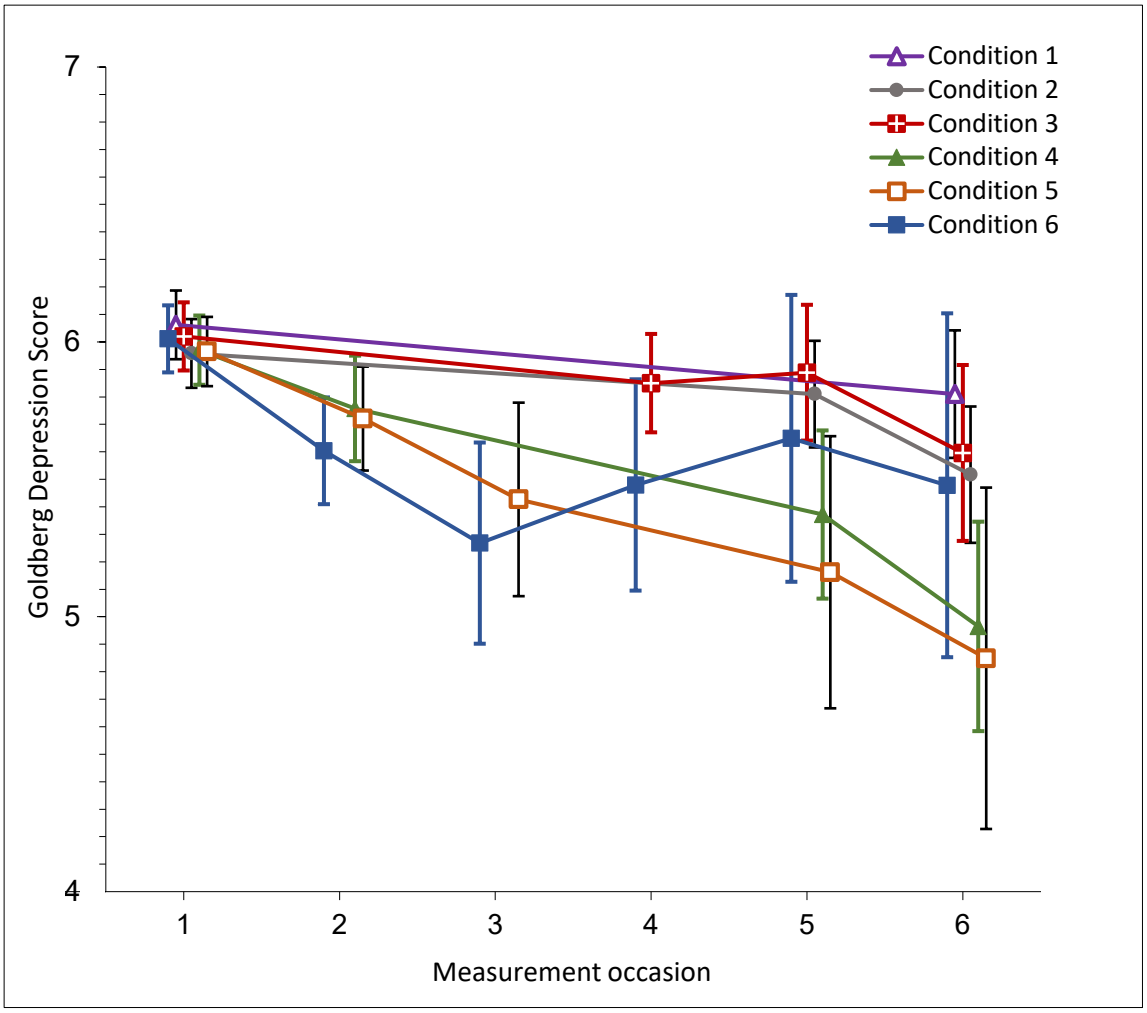

Figure 6.12. Estimated marginal means for Goldberg Depression score as a function of measurement occasion.

Between groups effect sizes were calculated for each of the conditions compared to Condition 1 (Module 1 only), using the difference between the estimated end-point means of the conditions and the pooled standard deviations of baseline scores. The effect sizes for Conditions 2 (Module 1 and 5), 3 (Modules 1, 4 and 5) and 6 (all modules) were negligible. The largest effect sizes compared to Condition 1 were for Condition 5 (Modules 1, 2, 3 and 5; $d = 0.31$) and for Condition 4 (Modules 1, 2 and 5; $d = 0.28$).

Contrasts were also calculated for changes from baseline to end-point scores between the conditions. The only conditions for which changes were significantly different compared to Condition 1 (Module 1 only) were Condition 4 (Modules 1, 2 and 5; $p < .001$) and Condition 5 (Modules 1, 2, 3 and 5; $p = .009$).

The baseline to end-point depression improvements were also significantly greater for the combination of conditions which included Module 2 (Conditions 4, 5 and 6) compared to those that

did not (Conditions 1, 2, 3; p = .004). In particular, Condition 4 (Modules 1, 2 and 5) resulted in significantly greater depression improvements than Condition 3 (Modules 1, 4 and 5; p = .016), and there were significantly greater improvements for Condition 4 (Modules 1, 2 and 5) compared to Condition 2 (Modules 1 and 5; p = .010). There were also significant differences for those conditions which included Module 5 (Conditions 2, 3, 4, 5, 6) compared to Condition 1, which did not (p = .002). However, the latter result includes all conditions compared to Condition 1, and as seen in the previous contrasts, depression changes were only significantly different for Conditions 4 and 5 compared to Condition 1. There was no difference between the combined conditions which included Modules 3 or 4 and those that did not.

6.4 Discussion

6.4.1 Summary of results

Aim 1a: Predicting choice or randomisation

Education was the most important predictor of choice status in a model which controlled for participant characteristics, with those with higher educational attainment more likely to choose their program variant. Older users were also more likely to choose, as were North American users compared to those from Oceania or Europe. Users were also more likely to choose if they had previously used anti-depressant medication, and increased anxiety symptom scores were associated with being more likely to choose. In contrast, higher baseline depression or self-reported history of depression were associated with being less likely to choose, after controlling for other variables.

Aim 1b: Adherence for those who chose and those who were randomised

A larger proportion of participants who were randomised completed their program (14.7%) compared to participants who chose their program (6.1%); however, this difference was strongly influenced by the fact that 70.4% of the latter group chose Condition 6 which comprised all five modules. Within the separate conditions, there was no difference in the proportion of users who completed the program for those who chose compared to those randomised, except for Condition 1 (Module 1 only). For this condition, a significantly larger percentage of randomised participants completed the program.

Adherence was also analysed based on whether participants completed at least two depression tests. Choosing a condition was a significant and important predictor of adherence on this measure after controlling for user characteristics and condition. Other significant predictors included higher

baseline Warpy Thoughts score, higher baseline depression score, younger age group, being located in Europe or Oceania compared to North America, and self-reported history of depression. After controlling for the other variables, including choice status, users who chose or were assigned to Condition 6 (all modules) were significantly more likely to complete two depression tests compared to those who chose or were assigned to Condition 1 (Module 1), Condition 3 (Modules 1, 4 and 5) and Condition 4 (Modules 1, 2 and 5).

Aim 1c: Depression outcomes for those who chose and those who were randomised

After controlling for baseline depression scores, there was a significantly greater improvement in depression score change for those users who chose their condition compared to those who were randomised. Choice status remained a significant predictor of depression improvement after controlling for participant's characteristics but did not significantly predict depression score change after controlling for the program variant that was chosen or received. In the final model increased depression improvement was significantly predicted by higher baseline depression, lower baseline anxiety and dysfunctional thinking, higher educational attainment, not having a self-reported history of depression, and undertaking Condition 6 (all modules) compared to Condition 1 (Module 1) or 3 (Modules 1, 4 and 5).

Aim 2: Comparisons of effectiveness for different program variants (RCT arm only)

Analysis of data from the RCT arm of the study suggested that participants randomised to Condition 1 (Module 1) had the smallest reduction of depression symptoms, although contrasts for Conditions 2, 3 and 6 compared to Condition 1 were not statistically significant, with negligible between-groups effect sizes. However, there were significant differences and small effect sizes, for Conditions 4 (Modules 1, 2 and 5, d = 0.28) and 5 (Modules 1, 2, 3 and 5, d = 0.31) compared to Condition 1. Conditions which included Module 2 resulted in significantly greater depression improvements than those that did not (p = .004). In particular Condition 4 (Modules 1, 2 and 5) was significantly superior to Condition 3 (Modules 1,4 and 5) and to Condition 2 (Modules 1 and 5).

6.4.2 Choice versus randomisation: Adherence and Outcomes

PRPTs are designed to accommodate the impact of participant preference for a specific treatment on user engagement with the intervention, or individual outcomes, particularly if a participant is otherwise randomised to an alternate condition. The current study is one of the first studies to investigate the effect of participant preference in depression iCBT interventions and provides the first evidence of which community users of Internet interventions are more likely to have a preference for a treatment option compared to those who are willing to be randomised.

It is noteworthy that a large majority of participants who chose their condition selected the full program. Whereas PRPTs usually allow for preferences between discrete treatment options (such as pharmaceutical or alternate psychological therapy approaches), the conditions in the current trial comprised combinations of the same modules. Hence it is likely that the participant's choice of condition was driven by a wish not to miss out on any part of the program, or a desire to decide for themselves which parts might be optimal.

In the current study, users were more likely to choose their program condition if they were older or more educated, or from North America compared to Oceania or Europe. It is possible that these users are more confident or self-empowered to make decisions about their own health based on their education, greater life experience or specific cultural background. Users with higher baseline anxiety scores were also more likely to choose their program. This may be associated with a heightened 'fear of missing out' on parts of the program should they agree to be randomised. Conversely, users with higher baseline depression scores were less likely to choose. Given that impaired decision-making capacity is associated with depression [199], it is possible that those with higher depression symptoms opted to simplify the decision-making process by agreeing to be assigned to an intervention rather than engaging in the more complex process of choosing between a number of program variants.

Alternatively, it is possible that users who were younger, less educated, less anxious, more depressed or from Europe or Oceania were more willing to fully participate in research than other users (i.e., that they were more motivated to contribute to scientific knowledge by being randomised, at the possible expense of missing out on a program for which they had a preference). However, there is no existing evidence base for this conclusion. There is a small amount of previous research on predictors of willingness to participate in mental health intervention research; however, a major difference between previous studies and the current study is that in the present research individuals were given the opportunity to participate by choosing their condition. Thus, refusal to be randomised was not a refusal to participate in research *per se*. Notwithstanding this difference, a recent study which examined the predictors of the willingness of outpatients with anxiety, mood and psychotic disorders to participate in mental health intervention research found that age and education were not related to willingness to participate [200]. Other previous studies have primarily focused on schizophrenia [e.g., 201] and Alzheimer's disease [e.g., 202] and so are not directly relevant to the current population.

It is likely that there are additional psycho-social or other variables which predict the likelihood that community users will opt to choose their intervention rather than accept randomisation. Possible

candidate variables include locus of control (level of self-belief in personal control over one's health), self-empowerment or self-efficacy. In addition, prior treatment with CBT is a possible candidate predictor given individuals with past CBT experience are likely to be more familiar with the components of CBT and thus potentially better placed to select preferred CBT components. Further research is needed to investigate whether such variables are associated with treatment preference. Finally, analysis of users' attitudes to participation in research trials may shed light on whether this influences the decision to choose or be randomised.

Choosing was a significant predictor of engagement with the program, as measured by completing at least two depression tests, after controlling for user characteristics and the condition. This supports existing evidence from research on face-to-face interventions that participant choice increases engagement [191, 192] and is the first evidence for such an effect in the context of iCBT interventions and depression iCBT programs in particular. In this light, the publicly available 'full' *MoodGYM* program may be too restrictive, requiring users to complete modules in order and with no option to skip or choose modules. However, choice was not a significant predictor of program *completion* after controlling for condition. This suggests that there is a limit to how much extra engagement with a self-directed iCBT program can be achieved by providing choice.

After controlling for user characteristics (but not condition), choosing was a significant predictor of depression score change. This is the first evidence that choosing a treatment option improves depression outcomes for iCBT depression intervention users. It is consistent with the results of a study of an open-access smoking cessation program which found improved outcomes for those who chose their web-based intervention components compared to those who were randomised to receive them as part of an RCT [193]. However, the results of the current study contrast with those of a study of a web-based physical activity intervention for which similar outcomes were found for participants whose randomly assigned delivery medium (text, video, or text and video) was matched to their preference compared to those where it was not [194]. The former study was most comparable to the current investigation since it allowed users to choose components of an existing online behavioural intervention (although the present study was more restrictive in the module combinations available to the user). It may be that choosing intervention content (as per the current study and the smoking cessation study) impacts on clinical outcomes, whereas choosing the delivery medium (as per the physical activity study) does not.

In the smoking cessation study [193] the analysis did not control for the components that were chosen by participants. However, in the current study, choice was no longer a significant predictor of depression score change after controlling for the condition that was received. In particular, Condition

1 (Module 1 only) and Condition 3 (Modules 1, 4 and 5) resulted in significantly less depression improvement than the full program, which suggests that users who received these variants missed out on one or more modules that may be important for increased effectiveness. One approach to optimising effectiveness might be to provide users with a core set of modules that are most critical for effectiveness (established by comparing the outcomes of different program variants using a randomised control design as per the second aim of the current study), followed by the opportunity to choose between different secondary modules or combinations of modules, to maximise engagement as per the previously discussed results. Future research could explore the effectiveness of this combined approach relative to full choice and no choice options. It might also explore which secondary modules work best for different users (e.g., based on age, gender, depression severity or engagement with the core modules), in order to inform a predictive recommendation algorithm. Based on these algorithms, future users could be assisted to make a choice of program modules via a recommendation (which they could choose to accept or ignore).

6.4.3 Which program variants are most effective?

Analysis of data from participants who were randomly assigned to the different program variants showed that the largest comparative reductions in depression symptoms were found for those assigned to Condition 4 (Modules 1, 2 and 5) and Condition 5 (Modules 1, 2, 3 and 5). These are the only two conditions (except for the full program) which included Module 2. There are some differences in these results compared to those previously reported for the *MoodGYM* variations trial (which included a sub-sample (n = 2,794) of the current dataset, n = 7,794) [118]. The previous study showed that all of the conditions except for Condition 6 (all modules) had significantly more improvement in depression symptoms compared to Condition 1 (Module 1 only) [118]; however, in the current study this was not the case for Conditions 2 (Modules 1 and 5) or 3 (Modules 1, 4 and 5). Hence the current study adds weight to the conclusion that Modules 1 and 2 (extended CBT) are the most important components for depression outcomes.

In both studies, contrasts between groups of conditions showed that those which included Module 2 resulted in superior outcomes compared to those that did not. However, unlike the previous study, in the current study Condition 4 (Modules 1, 2 and 5) was significantly superior to Condition 3 (Modules 1, 4 and 5), as well as Condition 2 (Modules 1 and 5).

As discussed in the previous publication [118], it is possible that differences in the lengths of the program conditions explain differences in depression outcomes – either because participants undertaking longer program variants had more time for their depression symptoms to resolve

spontaneously, or because they were exposed to more intervention content. However, neither explanation is convincing given that the most effective conditions were three and four modules long, and the less effective conditions included those which comprised an equal or increased number of modules.

There is also the possibility that Module 5 is an important intervention component. Indeed, contrasts between those conditions which included Module 5 and those that did not showed significantly greater improvements in depression symptoms for those which included Module 5. However, the only condition which did not include Module 5 was Condition 1, which was comprised solely of Module 1. Hence, additional research is required to better understand the importance of Module 5, which includes problem-solving techniques. In particular, research is required to compare Modules 1 and 2, with and without Module 5. For example, it would be instructive to take a multiphase optimisation strategy (MOST) approach to maximise the effectiveness of the intervention [203]. Within this framework, optimisation of module combinations could be investigated through a sequential, multiple assignment, randomised trial (SMART) – a type of factorial design where not all factors need to be crossed, and where multiple randomisations can be scheduled at different points in time [204]. This could build on the existing results by allowing for randomisation at specific points in the intervention to further identify key intervention components – for example, all users who completed Modules 1 and 2 could be randomised following the completion of Module 2 to either receive Module 5 or not. This, in turn, could inform the development of algorithms for guiding real-life users through the program.

Based on the current study, it appears that the 'extended CBT' modules (Module 1 and 2) may be core for effectiveness. However, there was a high attrition rate in all conditions, with only 41.6% of randomised participants who received the intervention completing the first module. Hence, as discussed in the previous section, given that allowing users to choose module combinations increases engagement, it may be possible to optimise both effectiveness and engagement of the program, by encouraging users to complete the core 'extended CBT' modules along with their choice of secondary module combinations.

6.4.4 Limitations

There are several limitations associated with the current study.

First, the use of a PRPT design complicates the interpretation and generalisation of the findings from the RCT arm of the trial. In particular, because this arm necessarily excluded individuals who had a preference for a particular condition, the sample that was randomised was reduced and hence was

not representative of the participants in the whole trial. Indeed, such complexity may be one reason that PRPTs are not more widely used. An alternative approach would have been to conduct an RCT with all participants but to measure treatment preference before randomisation, and to then examine the interaction between treatment preference and outcomes [205]. In addition, the treatment options presented to participants were brief with little explanation of how each option might assist an individual. Further information about each condition may have improved preference choices.

Second, although all randomised participants who completed the initial depression assessment were included in the analysis which compared outcomes for the different conditions, conclusions about the comparative effectiveness of different module combinations is limited by the small percentage of participants who completed their assigned modules (14.7%). In addition, mean improvements in depression symptoms were small, as were differences in improvement between conditions. However, although small changes may not represent clinical or meaningful change at an individual level, such changes and differences could translate into a benefit at a population level.

Third, the composition of the intervention conditions was not based on a fully factorial design, with some module combinations not included. This was partly a consequence of the need to limit the number of conditions in the design. Condition selection was undertaken based on hypotheses about which aspects of the program were likely to be most important (e.g. brief versus extended CBT), and was also limited by the need for some modules to follow on from the concepts taught in previous modules (e.g., Modules 2 and 3). However, the absence of combinations of conditions that permit the evaluation of the contribution of Module 5 (beyond the single module Condition 1) limits the conclusions that can be drawn from the RCT arm about the importance of this component. More recently, as noted above, researchers have investigated the optimisation of the effectiveness of multi-component behavioural interventions through fractional factorial designs within a MOST framework [e.g., 206]. Adoption of such an approach may have facilitated the identification and subsequent investigation of the importance (or not) of Module 5, without requiring a fully factorial trial design.

Participants in the study were self-selected, and their data were self-reported, with no external verification of the data that was entered. The outcome data (depression scale answers) were also collected in real time as part of the participants' use of the program. This means that some follow-up data were collected within a short time frame, and there was no standard period of time elapsed between baseline and follow-up.

Finally, as noted in the previous chapter, there was a risk of a 're-test artefact' in the current study. Thus, there is a possibility that administering the depression scale on multiple occasions led to a decrease in symptoms [184]. Given that different conditions included different numbers of depression assessments, it is possible that this re-test artefact confounded the results. However, it is of note in this context that the condition with the most depression assessments (Condition 6) was not found to be superior to shorter conditions.

6.5 Conclusion

The current study provides the first evidence of the characteristics of spontaneous, community users of iCBT who prefer to choose a program variant compared to those are willing to be randomised. Users who were older, had a higher level of education or who had more anxiety symptoms were more likely to choose, whereas users who had more depression symptoms were less likely to choose. The study also provides the first evidence that treatment choice may increase engagement with an iCBT intervention and adds to evidence from a previous trial [193] that user choice of program content improves outcomes in behavioural health Internet interventions. Hence, user choice may be an important aspect of optimising the effectiveness of iCBT interventions for some users and should be considered in the delivery of open access programs.

In addition, the current study extends a previous RCT of the relative effectiveness of different *MoodGYM* module combinations [118] and shows that the program variants comprising Modules 1, 2, 3 and 5, and Modules 1, 2 and 5, result in superior depression outcomes compared to other conditions. Contrasts between conditions suggest that Modules 1 and 2 are the core program components, and hence users should be encouraged to complete at least these two modules. Completion of at least two modules may also be a more meaningful adherence measure (and goal) than completion of the whole program. However, given that there was no condition which included only Modules 1 and 2, it is not clear whether or not Module 5 is an important component of the intervention. Further research is required to better understand how to maximise the effectiveness of the *MoodGYM* program for different community users. This exploration could be undertaken within a MOST framework, using a SMART trial design to investigate which and how modules should be delivered, beyond the first two core modules, for users with different characteristics.

7. Treatment moderators of MoodGYM

7.1 Introduction

The demographic and clinical characteristics which predict improved depression outcomes in spontaneous, community users of *MoodGYM* were examined in Chapter 5. Such variables are 'predictors' of outcome, also known as 'prognostic factors' [e.g., 207] and are important for identifying which users are most likely to improve over time. However, in order to identify factors which predict response to a particular treatment or intervention, a control (or comparison) condition is required, such as in an RCT design. In this context, treatment moderator variables (also known as 'treatment effect modifiers') are those baseline characteristics which affect the outcome for participants in the treatment or intervention condition, but not in the control condition [208]. In a model where a dependent variable is modelled over time, a moderator is indicated as a significant interaction between the moderator (independent) variable, condition and time [209] – that is, the impact of the moderator variable on the rate of change of the dependent variable is significantly different for participants in the treatment condition compared to those in the control condition.

Real-world observational studies of iCBT necessarily lack a control condition, and so investigations of moderators of iCBT are limited to controlled trials. However, there is an absence of RCTs which compare unguided iCBT to a control condition and which include moderator research. A possible reason for this is that the sample size required to generate sufficient power to detect the moderator interaction effects is approximately four times that required to detect overall treatment effects [210, 211]. Nonetheless, investigation of moderators, including exploratory or secondary analyses of RCT data, is important to understand better which users may benefit most from a particular intervention [212].

7.1.1 The literature on moderators and predictors of unguided iCBT for depression

The literature on moderators of unguided iCBT for depression is limited to a meta-analysis of individual participant data, conducted by Karyotaki and colleagues [28] and an RCT of the *Deprexis* program in a sample of participants with severe depression [98]. In the meta-analysis, data were analysed from 3,876 study individuals, sourced from 13 RCTs of self-guided iCBT compared to a control condition [28]. As noted in Chapter 2, the meta-analysis found that overall self-guided iCBT was effective in reducing depression symptoms (*SMD* = 0.27, 95% CI: 0.17, 0.37). In addition to calculating the overall treatment effect, possible moderators of treatment response were examined in the combined dataset. The analysis considered potential baseline demographic and clinical

moderator variables (age, sex, educational level, relationship status, employment status, comorbid anxiety and baseline depression severity) but found that iCBT treatment response was not moderated by any of these variables.

In the study of *Deprexis*, conducted with participants recruited from the community and health clinics who were experiencing severe depression symptoms (*N* = 396), Meyer et al. investigated whether a range of variables moderated depression in the unguided iCBT group but not in the control group [98]. Treatment response was found to be moderated by concurrent anti-depressant use – that is, participants in the iCBT condition who were taking anti-depressant medication improved more than those who were not whereas no such effect was found for the control condition. The analysis found that iCBT treatment response was not moderated by diagnosis with Major Depressive Disorder, baseline depression severity, age, sex, educational level, concurrent psychotherapy or age of onset of depression. However, given that the study was powered to detect overall treatment effects, it is possible that the sample size was too small to detect all moderator interaction effects. Data from this RCT was included in the Karyotaki meta-analysis [28].

A non-inferiority randomised trial comparing self-guided iCBT (*MoodGYM*), another self-guided iCBT program and self-guided Internet-delivered interpersonal psychotherapy (*N* = 1,843), also explored potential moderators of treatment response; however, this study did not incorporate a comparison to a control condition [213]. The study identified several baseline predictors of depression improvement across all conditions (being female, having a lower level of mastery or perceived control and having lower levels of dysfunctional thinking), but did not identify any moderators of outcome for the different treatments.

In view of the paucity of studies of moderators of unguided iCBT, it is useful to consider existing research on treatment modifiers of *guided* iCBT. Two RCTs of guided iCBT investigated moderators of treatment effect in comparison with a control condition, although both moderation analyses were limited by small sample sizes. In the first study, Button et al. conducted a secondary analysis of data from an RCT of guided iCBT for depression in a primary care setting in the United Kingdom (*N* = 297) [214]. They found that participants with higher levels of baseline depressive symptoms had a greater response to iCBT, and the iCBT treatment effect was also moderated by marital status, with larger improvement for those who received iCBT and who were separated, widowed or divorced, but that there was no effect for level of education, history of depression or number of stressful life events in the last 6 months (bereavement, separation or divorce, burglary or assault, court appearances, debt, disputes within close relationships, or redundancy).

In a Dutch RCT of 263 participants, which compared guided iCBT, guided Internet-delivered problem-solving therapy and a control condition, Warmerdam et al. found that depressive symptom improvement in all groups was predicted by higher baseline depression and higher levels of education; however, the only variable that moderated depression outcomes in the treatment groups compared to the control condition was 'negative problem orientation' (where individuals tend to be intimidated by problems and lack confidence in their ability to solve them) [215]. For this study, guidance was provided via email and was restricted to helping the user progress through the program.

In summary, there is some evidence that concurrent use of anti-depressant medication may moderate depression outcomes among users of *unguided* iCBT, and that depression severity, a 'negative problem orientation' and marital status may moderate depression outcomes among users of *guided* iCBT. However, no other baseline clinical, psychological or demographic variables have been found to moderate iCBT treatment response. Further, no controlled trial has investigated potential *moderators* of the *MoodGYM* intervention.

By contrast, several studies have reported significant *predictors* of outcomes among unguided users of iCBT. Christensen et al. identified female gender and initial anxiety level as predictors of depression outcome in 19,607 spontaneous users of *MoodGYM* [132]. As described in detail in Chapter 5 and summarised in tabular form in Table 7.4 below, in a more comprehensive study of 106,882 spontaneous *MoodGYM* users, the current candidate identified five demographic (female gender, older age group, higher educational level, being located in Australia rather than the United Kingdom, rural location), three clinical (baseline depression, anxiety, dysfunctional thinking), and five health or psychological-related (self-reported depression history, previous anti-depressant medication use, experience with CBT, number of previous depression treatments, motivation) predictors of depression outcome. In a much smaller study, comprising older individuals with sub-threshold depression who received either unguided iCBT (*n* = 102) or group CBT (*n* = 99), Spek and colleagues found that higher baseline depression severity, higher educational level, being female, and lower neuroticism predicted depression improvement among all users although other personality factors (extraversion, openness, altruism and conscientiousness), previous episodes of depression, and marital status were not significant predictors of depression outcome [216]. In contrast to the above studies, Berger et al. failed to identify any significant predictors of depression outcome in a trial of the depression iCBT program *Deprexis.* In particular, depression outcome was not predicted by gender, age, education, previous treatment, medication or co-morbid mental health

conditions [104]. However, this study incorporated the combined the results of guided and non-guided CBT, and the total sample size was small (*n* = 69).

7.1.2 Current study

The current study extends the limited existing research into moderators of iCBT treatment by exploring possible moderator variables in an automated RCT of *MoodGYM* (unguided) compared to a wait-list control. This study is a secondary analysis of data obtained from the Psywell study, which evaluated the use of *MoodGYM* to increase mental wellbeing in visitors to the NHS Choices health information website [103]. The candidate was involved in the planning, set-up and trial management of the Psywell study and was a co-author on the published paper outlining the primary results. The current study was undertaken by the candidate independently of the original investigation with the guidance of the candidate's supervisory panel.

Data from the Psywell study were not eligible for inclusion in the Karyotaki meta-analysis of unguided iCBT for depression [28] since the primary outcome measure for the study was mental wellbeing measured using the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS), and not depression. Depression was a secondary outcome for the trial, measured using the CES-D.

The RCT involved 3,070 adult participants located in England, of whom approximately half (49.8%) completed follow-up assessments after 12 weeks. Participants randomised to the *MoodGYM* condition completed, on average, 1.3 of the five modules, and one third (33.8%) completed two or more modules. The published paper included an intent-to-treatment analysis of depression outcomes and found that depression was significantly reduced in the *MoodGYM* group compared to the wait-list control, with a highly significant ($p < .001$) time by condition interaction at both 6- and 12-weeks [103]. There was no difference in CES-D scores between the intervention and control groups at baseline (difference: -0.04, 95% CI: -0.99, 0.91) and significant improvements for the *MoodGYM* group at 6 weeks (difference: -2.79, 95% CI: -3.95, -1.64) and at 12 weeks (difference: -3.36, 95% CI: -4.62, -2.11). However, no sub-group analyses were conducted.

This data provides an opportunity to identify potential moderators of the effectiveness of *MoodGYM* for depression, in a setting approximating that of spontaneous, community users. Although the data is from individuals recruited to a research trial (and hence, as discussed in Chapter 4, the results may not translate directly into a real-world context), the differences introduced by the research protocol were minimal. In particular, participants were recruited from a health information website (and so were already seeking help online), and inclusion criteria were as broad as possible. The trial process was also fully automated, with no human contact. Other than the knowledge that they were

participating in a trial and could be allocated to a wait-list condition, the main difference between the experience of the Psywell participants and that of spontaneous, community users was that the study participants in the intervention condition received weekly automated emails in the intervention period encouraging them to log in to access the program.

Study Aim

The primary aim of the current study is to investigate whether demographic or baseline clinical characteristics *moderated depression outcomes* for participants in the iCBT condition – i.e., whether there are any participant characteristics which affect the magnitude of depression improvement for those in the group that was randomised to receive *MoodGYM* but not for those in the wait-list control. In order to consider the data in the context of earlier chapters that investigated the *predictors of depression outcomes* among spontaneous users of *MoodGYM*, the secondary aim of the current study is to separately identify the predictors of depression improvement across both the intervention and control conditions.

7.2 Method

The study received approval from the Australian National University Human Research Ethics Committee (Protocol Number 2010/244) and from the NHS ethics committee (Black County REC 10/H1202/21) and was registered with the International Standard Randomised Controlled Trial Number Register (ISRCTN 48134476).

7.2.1 Participants

Participants were recruited through the Live Well and mental health pages of the NHS Choices website over two weeks in September 2010. The advertisement offered individuals an opportunity to take part in a mental fitness trial to promote wellbeing. The study did not mention depression or other mental health issues and was not described as involving treatment or therapy.

Inclusion criteria

Participants were eligible for participation if they:

- were aged 18 years or older;
- were located in England;
- were able to read and write English;
- had access to the Internet; and

- had an email address where they could be contacted.

Exclusion criteria

There were no additional exclusion criteria.

7.2.2 Procedure

Consent and randomisation

After screening against the inclusion criteria, eligible participants were instructed to create a username and password and were then sent an email 48 hours later, inviting them to provide informed consent to participate in the trial. This delay was included to allow potential participants to reflect on their decision to take part in the trial. Once participants had provided consent, they were directed to the baseline demographic and health survey. After submission of the baseline survey, participants were randomised via an automated computer algorithm to either the intervention group or the control group. Randomisation was conducted on a 1:1 basis with a block size of two.

Trial procedure

Participants in the intervention group were provided with access to the five modules of the *MoodGYM* program, which they were able to work through at their own pace (as per the publicly available version of the program). The program was near identical to the open-access program, except that participants accessed the program through a study portal (which was branded with the NHS and University of Warwick logos), and some minor content modifications to the program were made to ensure that the program was suitable for U.K. audiences (for example replacing Australian colloquialisms). At 1, 3 and 5 weeks after randomisation, participants in the intervention condition received an automated email reminding them to log in to the program, irrespective of whether they had previously accessed the intervention. Participants who were randomised to the control group were informed that they would be given access to the intervention in three months and were directed to the Healthy Living webpages of the NHS Choices website. To avoid cross-contamination, the name of the program was not included in any of the study materials.

At 6 weeks and 12 weeks after randomisation, all participants received an email invitation to complete follow-up surveys, with up to two reminder emails. After 12 weeks, participants in the control group were also provided with access to the intervention through the study portal.

7.2.3 Measures

Primary outcome measure

Participants' depression severity was measured at baseline and 6- and 12-weeks follow-up using the Centre for Epidemiologic Studies Depression Scale (CES-D) [217], a 20 item depression inventory which has been validated for use via the Internet [218]. The scale asks about symptoms over the past week, with each item scored from 0 to 3. The total score range for the CES-D is 0 to 60, with higher scores representing increased depressive symptoms. A score of 16 or higher is widely used as a clinical cut-off score for depression [e.g., 219].

Other measures

The following measures were administered at baseline and 6- and 12-week follow-up:

Anxiety: The GAD-7 survey, which has been validated for use in Internet settings, was used to measure anxiety (7 items asking about symptoms in the last two weeks, with responses on a 4-point scale) [220]. The GAD-7 is scored from 0 to 21 with higher scores reflecting higher anxiety severity. Mild, moderate and severe anxiety symptoms are reflected at cut-off scores of 5, 10 and 15, respectively.

Wellbeing: Wellbeing was measured using the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) – 14 items asking about feelings and thoughts over the past two weeks, with answers on a 5-point scale [221]. The scale is scored from 14 to 70, with increased scores represent higher levels of mental wellbeing. This measure was the primary outcome measure for the original study [103].

Quality of Life: EuroQol Group 5-Dimension Self Report Questionnaire (EQ-5D) was used to measure health status and quality of life [222]. The survey asks about five health dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) and participants are also required to rate their health state on a scale of 0 to 100 (with 100 being the best possible state and 0 being the worst state).

Physical activity: A single item, validated measure of physical activity asked participants to indicate on how many days in the past week they had undertaken a total of 30 minutes or more of physical activity [223].

In addition, questions about the following demographic and health-related variables were included in the baseline survey only:

Demographic characteristics: Participants were asked to indicate their sex, age, marital status, employment status and level of educational qualifications. The wording and response categories for these questions were taken directly from the 2011 Census for England and Wales rehearsal questionnaire [224]. Participants were also asked to report how often they used the Internet.

Health-related variables: Participants were asked if they had ever received treatment for a mental health problem, if they had ever previously had CBT, and if they had ever used an Internet intervention for wellbeing, depression or anxiety.

7.2.4 Analyses

Potential moderators of depression outcome were coded as follows:

- Sex: [Male / Female];
- Age: Years from 18 to 100;
- Marital status: [Married / Divorced or separated / Never married / Widowed];
    - Coded 'Married' if 'Married' or 'In a registered same-sex civil partnership' selected.
    - Coded 'Divorced or separated' if 'Separated but still legally married', 'Divorced', 'Separated but still legally in a same-sex civil partnership' or 'Formerly in a same-sex civil partnership which is now legally dissolved' selected.
    - Coded 'Never Married' if 'Never married and never registered in a same-sex civil partnership' selected.
    - Coded 'Widowed' if 'Widowed' or 'Surviving partner from a same-civil partnership' selected.
- Employment status: [Working / Student / Retired / Carer / Unemployed];
    - Coded 'Working' if response to 'Which of these descriptions best applies to what you were doing last week?' was 'In paid employment (full or part-time) or self-employed (or temporarily away)' or 'Doing unpaid work for a business that you own, or that a relative owns'.
    - Coded 'Student' if 'Going to school or college full-time (including on vacation)' selected.
    - Coded 'Retired' if 'Retired from paid work' selected.
    - Coded 'Carer' if 'On maternity leave' or 'Looking after family or home' selected.

- Coded ‘Unemployed’ if ‘On a Government scheme for employment training’ or ‘Unemployed’ or ‘Permanently unable to work because of long-term sickness or disability’ or ‘Doing something else’ selected.

- Educational qualifications: [None, or primary / Secondary / Tertiary];
  - Coded ‘Tertiary’ if selections in response to ‘Which of these qualifications do you have?’ included ‘Degree (for example BA, BSc), Higher degree (for example MA, PhD, PGCE)’.
  - Coded ‘Secondary’ if not classified as Tertiary, and selections included ‘5+ O levels (passes)/CSEs (grade 1)/GCSEs (grades A*-C), School Certificate, 1 A level/2-3 AS levels/VCEs, Higher Diploma’.
  - Coded ‘None, or primary’ if not classified as Tertiary or Secondary.
- Smoking status: [0, 1];
  - Coded 0 if response to ‘How often do you currently smoke?’ was ‘Not at all’.
  - Coded 1 if ‘Daily’ or ‘Occasionally’ selected.
- Daily Internet use: [0, 1];
  - Coded 1 if response to ‘In general, how often do you use the Internet?’ was ‘At least once a day’.
  - Coded 0 otherwise.
- Previous mental health treatment: [0, 1];
  - Coded 0 if response to ‘Have you ever received treatment for a mental health problem that you consulted a doctor about?’ was ‘No’.
  - Coded 1 if ‘Yes’ selected.
- Previous CBT: [0, 1];
  - Coded 0 if response to ‘Have you ever had cognitive behavioural therapy?’ was ‘No’.
  - Coded 1 if ‘Yes’ selected.
- Previous Internet intervention: [0, 1];
  - Coded 0 if responses to ‘Have you ever used a website training programme for mental wellbeing?’ and ‘Have you ever used a website training programme to treat depression or anxiety?’ were ‘No’.

- Coded 1 if 'Yes' selected for either question.

- GAD-7 (baseline): Score from 0 to 21;
- WEBWBS (baseline): Score from 14 to 70;
- Physical activity (baseline): Score from 0 to 7.

Each of the five dimensions of the EQ-5D at baseline was considered as a separate potential moderator variable, and responses for each were dichotomised:

- Mobility problems: [0, 1];
    - Coded 0 if 'I have no problems in walking about' selected.
    - Coded 1 if 'I have some problems in walking about' or 'I am confined to bed' selected.
- Self-care problems: [0, 1];
    - Coded 0 if 'I have no problems with self-care' selected.
    - Coded 1 if 'I have some problems washing or dressing myself' or 'I am unable to wash or dress myself' selected.
- Usual activities problems: [0, 1];
    - Coded 0 if 'I have no problems with performing my usual activities' selected.
    - Coded 1 if 'I have some problems with performing my usual activities' or 'I am unable to perform my usual activities' selected.
- Pain problems: [0, 1];
    - Coded 0 if 'I have no pain or discomfort' selected.
    - Coded 1 if 'I have moderate pain or discomfort' or 'I have extreme pain or discomfort' selected.
- Self-reported anxiety/depression: [0, 1];
    - Coded 0 if 'I am not anxious or depressed' selected.
    - Coded 1 if 'I am moderated anxious or depressed' or 'I am extremely anxious or depressed' selected.

The EQ-5D health state response was also considered as a separate variable:

- Health state (baseline): Score from 0 to 100.

The demographic, health and clinical characteristics of participants in the two conditions were examined, and Chi-square and ANOVA tests performed to check for any significant differences between the groups.

Linear mixed effects models for repeated measures were then constructed to identify which variables were predictors of depression outcome (across both two conditions) or moderators of depression outcome in response to treatment via iCBT. Mixed models were used since this enables an ITT analysis, using data from all participants (not only those who completed follow-up assessments). The analysis assumed that missing observations were 'missing at random' (MAR) and so may be related to the participant or intervention characteristics, but not to the value of the missing data. As discussed in previous chapters, in intervention effectiveness research this approach is considered superior to other non-ITT analysis methods [122]. The method for identifying predictors and moderators mirrored that employed by Donker and colleagues in their study of predictors and moderators of response to iCBT and Internet-delivered interpersonal psychotherapy [213].

Predictors of improvement in depression outcome

A linear mixed model for repeated measures analysis was conducted with CES-D as the dependent variable, measured at baseline, 6-week post-test and 12-week follow-up. Fixed effects were condition (intervention or control) and the potential baseline predictors of depression outcome, as well as the interaction of each with time. The model used restricted maximum likelihood estimation and an unstructured variance covariance structure. Type III omnibus tests of fixed effects were examined to identify statistically significant fixed effects and estimated marginal means were calculated to explore the direction of the association between predictors and CES-D scores.

Moderators of iCBT treatment effect

Separate linear mixed models for repeated measures were constructed to explore whether each of the baseline variables moderated improvement in depression outcome for those in the iCBT condition. Each model used CES-D as the dependent, repeated variable, and fixed effects were the potential moderator, time, condition, all two-way interactions, and the three-way interaction between the moderator, time and condition. Type III omnibus tests were examined for the fixed effects to determine if any of the three-way interaction effects were statistically significant, which

would indicate that the investigated variable moderated depression outcomes over time for those in the iCBT condition compared to the control condition.

As per the method used by Donker et al., each potential moderator was considered separately due to the large number of interaction terms that would need to be included if all potential moderators were included in the same model. There was no correction for multiple analyses.

All statistical analyses were conducted using SPSS Version 25.0 [140].

7.3 Results

The demographic, health and clinical characteristics of users are displayed in Table 7.1, grouped by condition. Chi-square tests revealed no significant differences between the groups for any of the categorical variables, and similarly, ANOVA tests showed no significant differences in the means of continuous variables between the intervention and control conditions.

Table 7.1. Baseline survey characteristics of participants, by condition (n=3,070).

	MoodGYM *n* **(%)**	**Control** *n* **(%)**	**Total** *n* **(%)**
Total	1,534	1,536	3,070
Sex			
Female	1,179 (76.9%)	1,212 (78.9%)	2,391 (77.9%)
Male	355 (23.1%)	324 (21.1%)	679 (22.1%)
Marital status			
Married / co-habiting	722 (47.1%)	711 (46.7%)	1,433 (46.7%)
Divorced / separated	246 (16.0%)	252 (16.4%)	498 (16.2%)
Never married	545 (35.5%)	533 (34.7%)	1,078 (35.1%)
Widowed	21 (1.4%)	40 (2.6%)	61 (2.0%)
Employment status			
Working	953 (62.1%)	916 (59.6%)	1,869 (60.9%)
Student	56 (3.7%)	61 (4.0%)	117 (3.8%)
Retired	93 (6.1%)	101 (6.6%)	194 (6.3%)
Carer	202 (13.2%)	234 (15.2%)	436 (14.2%)
Unemployed	230 (15.0%)	224 (14.6%)	454 (14.8%)
Educational qualification			
Primary or none	530 (34.6%)	562 (36.6%)	1,092 (35.6%)
Secondary	372 (24.3%)	391 (25.5%)	763 (24.9%)
Tertiary	632 (41.2%)	583 (38.0%)	1,215 (39.6%)
Smoker			
False	1,213 (79.1%)	1,214 (79.0%)	2,427 (79.1%)
True	321 (20.9%)	322 (21.0%)	643 (20.9%)
Daily Internet use			
False	173 (11.3%)	188 (12.2%)	361 (11.8%)
True	1,361 (88.7%)	1,348 (87.8%)	2,709 (88.2%)
Previous mental health treatment			
False	657 (42.8%)	693 (45.1%)	1,350 (44.0%)
True	877 (57.2%)	843 (54.9%)	1,720 (56.0%)
Previous CBT			
False	1,208 (78.7%)	1,215 (79.1%)	2,423 (78.9%)
True	326 (21.3%)	321 (20.9%)	647 (21.1%)
Previous use of Internet interventions			
False	1,395 (90.9%)	1,402 (91.3%)	2,797 (91.1%)
True	139 (9.1%)	134 (8.7%)	273 (8.9%)

Table 7.1. Baseline survey characteristics of participants, by condition (n=3,070) (continued).

	MoodGYM **n *(%)***	***Control*** **n *(%)***	***Total*** **n *(%)***
Problems with mobility			
False	1,282 (83.6%)	1,273 (82.9%)	2,555 (83.2%)
True	252 (16.4%)	263 (17.1%)	515 (16.8%)
Problems with self-care			
False	1,437 (93.7%)	1,425 (92.8%)	2,862 (93.2%)
True	97 (6.3%)	111 (7.2%)	208 (6.8%)
Problems with usual activities			
False	1,079 (70.3%)	1,090 (71.0%)	2,169 (70.7%)
True	455 (29.7%)	446 (29.0%)	901 (29.3%)
Problems with pain or discomfort			
False	861 (56.1%)	859 (55.9%)	1,720 (56.0%)
True	673 (43.9%)	677 (44.1%)	1,350 (44.0%)
Self-reported anxiety or depression			
False	560 (36.5%)	559 (36.4%)	1,119 (36.4%)
True	974 (63.5%)	977 (63.6%)	1,951 (63.6%)
	MoodGYM ***M* (*SD*)**	**Control** ***M* (*SD*)**	**Total** ***M* (*SD*)**
Age (years)	40.88 (12.95)	41.39 (13.05)	41.14 (13.01)
CES-D	23.23 (13.26)	23.27 (13.65)	23.25 (13.45)
WEBWBS	42.20 (9.82)	42.32 (9.66)	42.26 (9.74)
GAD-7	8.80 (5.93)	8.46 (5.82)	8.63 (5.88)
Health state (0 to 100)	69.12 (21.10)	69.04 (20.40)	69.08 (20.75)
Physical activity sessions	2.17 (2.01)	2.23 (2.06)	2.20 (2.04)

The mean age of participants was 41.1 years (*SD* = 13.01) and most were female (77.9%). Nearly half were married or living with a partner (46.7%), and a majority were currently employed (60.9%). Approximately 40% of participants had a university degree and most reported daily Internet use (88.2%). More than half had previously sought treatment for a mental health issue (56%) although only 21.1% had tried CBT. A majority self-reported having depression or anxiety, which was reflected in the mean CES-D (23.25, *SD* = 13.45) and GAD-7 (8.63, *SD* = 5.88) scores at baseline.

7.3.1 Predictors of improvement in depression outcome

A multivariate linear mixed model for repeated measures was constructed to investigate if any of the demographic and baseline clinical variables predicted improvement in CES-D scores at post-test and

12-week follow-up. All potential predictor variables and an interaction term between each variable and time were entered simultaneously as fixed effects in the model, which used an unstructured covariance structure (-2 Restricted Log Likelihood = 42,398.48). The results of Type III omnibus tests of the fixed effects are displayed in Table 7.2.

Table 7.2. Potential predictors of CES-D scores (n=3,070).
Estimates are Type III omnibus tests from a multivariate linear mixed model for repeated measures.

Source	F	df	*p*
Intercept	795.6	2201.7	<.001
Time	18.8	1709.0	**<.001**
Condition	59.7	2296.0	**<.001**
Sex	2.3	2182.7	.128
Marital status	5.6	2149.8	**.001**
Employment status	2.6	2172.5	**.032**
Educational qualification	2.9	2170.7	.058
Smoking status	4.2	2213.5	**.040**
Daily Internet use	0.0	2203.1	.876
Previous mental health treatment	14.2	2168.3	**<.001**
Previous CBT	6.5	2180.9	**.011**
Previous use of Internet interventions	0.3	2208.7	.598
Problems with mobility	0.4	2184.4	.546
Problems with self-care	2.9	2169.6	.090
Problems with usual activities	8.5	2194.6	**.004**
Problems with pain or discomfort	2.0	2176.6	.160
Self-reported anxiety or depression	11.7	2173.6	**.001**
Age (years)	2.8	2169.2	.092
WEBWBS	557.3	2181.0	**<.001**
GAD-7	338.0	2210.8	**<.001**
Health state (0 to 100)	22.1	2190.7	**<.001**
Physical activity sessions	1.2	2137.4	.278
Time interaction terms			
Condition	24.8	1722.6	**<.001**
Sex	0.4	1700.1	.691
Marital status	2.5	1704.5	**.023**
Employment status	1.8	1705.6	.071
Educational qualification	1.1	1700.3	.337
Smoking status	1.0	1700.8	.379
Daily Internet use	0.3	1707.8	.720

Table 7.2. Potential predictors of CES-D scores (n=3,070) (continued).

Source	*F*	*df*	p
Previous mental health treatment	5.1	1702.0	**.006**
Previous CBT	4.1	1706.0	**.017**
Previous use of Internet interventions	0.1	1742.1	.878
Problems with mobility	2.2	1712.8	.117
Problems with self-care	0.0	1709.4	.979
Problems with usual activities	0.2	1707.6	.788
Problems with pain or discomfort	1.7	1702.2	.181
Self-reported anxiety or depression	2.4	1702.6	.090
Age (years)	4.1	1703.7	**.016**
WEBWBS	31.8	1707.4	**<.001**
GAD-7	5.9	1715.2	**.003**
Health state (0 to 100)	0.2	1699.4	.815
Physical activity sessions	1.3	1699.4	.283

Overall, depression scores were significantly higher for participants who were widowed or never married and were lowest for participants who were married. Participants who were employed or who were students had significantly lower scores than those who were retired, unemployed or caring for family. Smokers and participants who had previously accessed mental health treatments or CBT had significantly higher depression scores overall, as did those participants who had problems with their usual activities or self-reported having depression or anxiety. Lower health state and wellbeing scores and higher baseline anxiety were also associated with higher CES-D scores.

The interactions of demographic and baseline clinical variables with time in the model suggests that there are a number of predictors of outcome across both conditions. Marital status was a significant predictor of improvement in depression symptoms in participants, with those who were divorced improving most after controlling for the other variables, whereas participants who were widowed or never married improved the least. Participants who had previously accessed treatment for mental health issues or who had previously experienced CBT improved significantly less than those who had not. In addition, being younger, having higher baseline anxiety and having lower baseline mental wellbeing all predicted greater depression improvement at post-test and follow-up.

7.3.3 Moderators of treatment

The results of separate repeated measures linear mixed models for potential treatment moderators are shown in Table 7.3. Only one of the models had a significant three-way interaction term; problems with self-care (washing/dressing oneself) moderated depressive symptom outcomes after treatment with *MoodGYM, F*(2, 1,792.4) = 3.01, *p* = .05. Specifically, participants who did not have issues with self-care experienced improved depression outcomes after the iCBT intervention, whereas those with self-care issues did not.

Table 7.3. Three-way interactions of time x condition x potential moderator, for potential treatment moderators of CES-D scores (n=3,070).
Estimates are Type III omnibus tests from separate repeated measures linear mixed models analyses.

Source	*F*	df	*p*
Time x condition x potential moderator interaction terms			
Sex	1.2	1770.5	.311
Marital status	0.8	1811.5	.559
Employment status	0.3	1799.3	.962
Educational qualification	0.4	1763.5	.814
Smoking status	0.6	1771.5	.636
Daily Internet use	2.5	1778.0	.080
Previous mental health treatment	1.0	1772.6	.357
Previous CBT	0.2	1775.9	.835
Previous use of Internet interventions	0.2	1789.3	.827
Problems with mobility	1.0	1802.9	.373
Problems with self-care	3.0	1782.4	**.050**
Problems with usual activities	0.0	1776.8	.965
Problems with pain or discomfort	0.2	1770.3	.822
Self-reported anxiety or depression	1.6	1755.1	.194
Age (years)	0.2	1769.0	.783
WEBWBS	1.1	1731.1	.320
GAD-7	1.2	1750.1	.294
Health state (0 to 100)	0.4	1766.3	.660
Physical activity sessions	1.4	1774.8	.254

The estimated marginal means for participants in the control and iCBT conditions, grouped by those who reported that they did and did not have problems with self-care, are shown in Figure 7.1. Visual inspection confirms that the rate of improvement in depression symptoms is greater in the *MoodGYM* group than the control group for participants with no self-care problems. However, there

is no difference in improvement between conditions for participants who do have issues with self-care. Hence, depression outcome is moderated by self-care status.

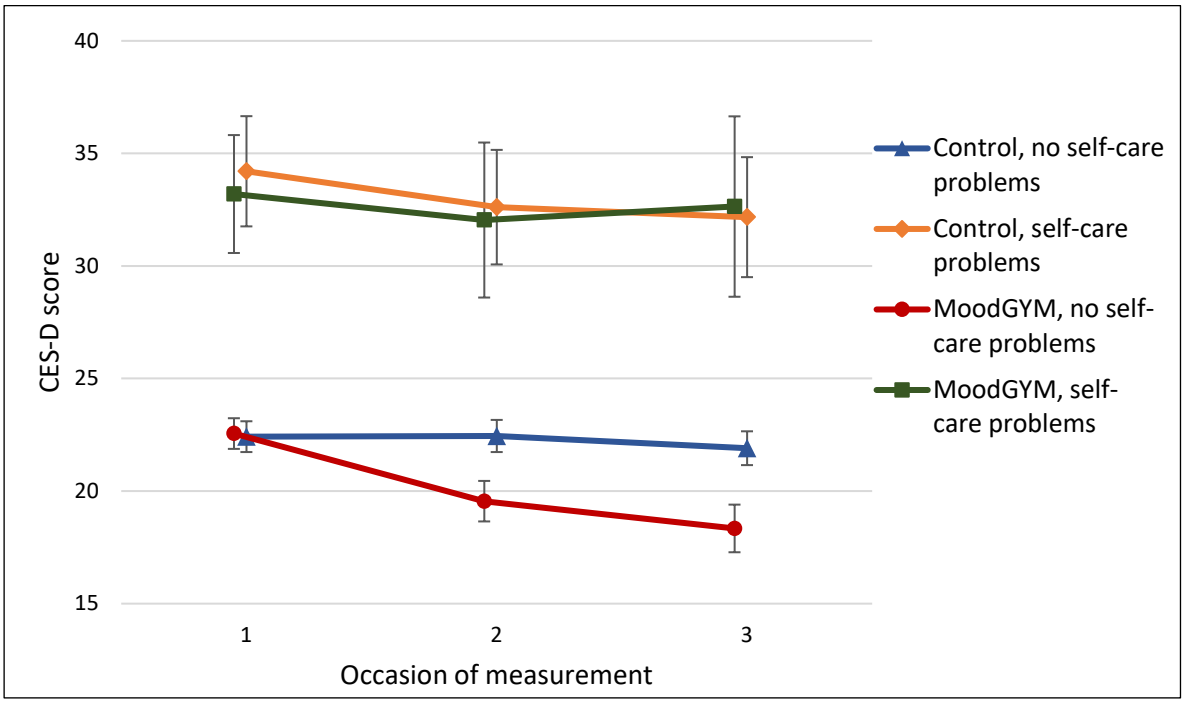

Figure 7.1. Problems with self-care as a moderator of depressive symptoms after iCBT. Estimated marginal means from repeated measures linear mixed model analysis, error bars represent 95% confidence intervals.

7.4 Discussion

7.4.1 Summary of results

Several variables predicted improvement on the CES-D in participants in both conditions. Participants who were divorced improved more than those who were married, whereas those who were never married or who were widowed improved least. Not having accessed either mental health treatment or CBT previously predicted larger improvement in depression scores. Greater improvement was also predicted by being younger, having higher baseline anxiety and having lower mental wellbeing.

Exploratory analyses found that depression outcomes after iCBT were moderated by only one of the investigated baseline demographic and clinical participant characteristics. Participants who reported having trouble with washing or dressing did not improve after iCBT, whereas those who did not had improved outcomes. This difference was not found in the wait-list control group.

7.4.2 Predictors of depression outcome

One of the two aims of the current study was to compare the predictors identified in the current RCT of the *MoodGYM* intervention with the predictors of outcome among spontaneous *MoodGYM* users as reported in Chapter 5. In fact, there were limited similarities between the predictors of depression outcome in the current dataset (both conditions) compared to those found in the much larger sample of spontaneous, community users of *MoodGYM* (reported in Chapter 5). Table 7.4 displays the demographic, health-related and clinical variables investigated in each of the predictor analyses[7]. The two studies shared seven potential predictor variables in common: sex, age or age group, educational level, previous use of CBT, self-reported depression (or anxiety), previous mental health treatment/use of anti-depressant medication and initial anxiety. Not having accessed mental health treatment (current study) and not having used anti-depressant medication (spontaneous users study) both predicted larger improvement. However, whereas *younger* participants and those with *higher* baseline anxiety improved more in the current study, older users and those with lower anxiety improved more in the spontaneous users study. In addition, self-reported depression, being female and higher education were significant predictors of depression outcome in the spontaneous users study but did not predict improvement in the current study. Conversely, having previously tried CBT was a significant predictor of less improvement in the current study only. Marital status and mental wellbeing (which were predictors of outcome in the current study) were not measured in the previous study.

[7] The relevant analysis from Chapter 5 is Step 2 of the multivariate regression reported in Table 5.13. Step 3 of the analysis is not comparable to the current study since it includes adherence variables related to use of the intervention. The dataset in the current study comprises both intervention and control conditions, and hence the predictor analysis is limited to baseline characteristics of participants.

Table 7.4. Potential predictor variables considered in models of depression outcome.

✓ = significant predictor; x = non-significant variable; shaded cells represent variables that were not investigated in the study.

		Current study (N=3,070)	Chapter 5 Spontaneous user study (n=106,882)
Demographic variables	Sex	x	✓
	Age[a]	✓	✓
	Educational level	x	✓
	Marital status	✓	
	Employment status	x	
	Daily Internet use	x	
	Country		✓
	Rural location		✓
Health-related variables	Previous use of CBT	✓	x
	Self-reported depression/anxiety[b]	x	✓
	Previous mental health treatment[c]	✓	✓
	Smoking status	x	
	Previous use of Internet intervention	x	
	Level of physical activity	x	
	Mobility problems	x	
	Self-care problems	x	
	Usual activities problems	x	
	Health state (0 to 100)	x	
	Referral by a health professional		x
	Motivation level		✓
	Experience/knowledge of CBT		✓
	Number of previous treatments		✓
Clinical variables	Anxiety[d]	✓	✓
	Mental wellbeing (WEBWBS)	✓	
	Baseline depression (Goldberg)		✓
	Warpy Thoughts score		✓

[a] *Current study: Age; Spontaneous user study: Age group*

[b] *Current study: Self-reported depression or anxiety; Chapter 5 study: self-reported depression only.*

[c] *Current study: Previous mental health treatment; Chapter 5 study: Previous use of anti-depressants.*

[d] *Current study: GAD-7; Chapter 5 study: Goldberg Anxiety scale.*

It is possible that the differences in findings for the two studies signify that the predictors of outcome differ for research compared to spontaneous participants. This would suggest that caution be exercised in generalising from the results of research trials to real-world contexts. However, it is also possible that the differences in the results of the two studies were due to key methodological divergences. First, the sample size for the spontaneous users study (*n* = 106,882) was approximately 35 times larger than the current study, and thus, the former had higher power to detect predictors. Second, the spontaneous users study included only those users with at least two depression survey responses whereas the intent to treat analysis employed in the current study included all participants (including those with no follow-up observations). Hence, whereas the current predictor analysis sought to determine which participant characteristics predicted depression improvement regardless of condition, the previous study considered only those users who had at least engaged with the first module of the program, potentially resulting in a biased sample. In addition, the studies used different measures of depression and anxiety (CES-D and Goldberg Depression Scale; GAD-7 and Goldberg Anxiety Scale), and follow-up surveys were administered at fixed timepoints in the current study but as part of the intervention in the previous study. Finally, there were demographic differences in the trial samples which may have contributed to differences in the results. For example, participants in the current study were located in England, whereas the spontaneous users study included users from the United Kingdom, Canada, United States and Australia. In the previous study being located from the U.K. *predicted* decreased depression improvement. The current study also required participants to be aged 18 years or older, whereas there was no age restriction in the spontaneous users study, with 9.4% being aged 19 years or younger. Overall it appears that participants in the current study were older: the median age of participants was 40 years; but only 34.9% of the spontaneous users sample was aged 40 years or older.

In order to better understand which factors predict improvement in individuals who are actively seeking health information online, further research is required – in particular, similar RCTs to the current study but with additional potential predictor variables, for example, personality factors and level of dysfunctional thinking. However, from the perspective of determining who will most benefit from an online intervention, it is a consideration of the findings of the moderation analysis (below) which is critical.

7.4.3 Self-care as a moderator of depression outcome

The findings of the current study broadly confirm those of the Karyotaki et al. study of the moderators of depression outcome in unguided depression iCBT interventions, which, using an individual participant data meta-analysis [28], found no demographic or clinical moderators of

outcome. However, since the use of anti-depressant medication was not assessed in the current study, it is not possible to compare the current findings with those reported by Meyer et al. in their study of unguided *Deprexis* [98]. As can be seen in Table 7.5, which lists the variables considered as potential moderators in each study, the current study extended the previous investigations by demonstrating null findings for a range of additional variables considered as potential treatment effect modifiers. However, significantly, the current study also yielded a novel finding, demonstrating that an absence of problems with washing or dressing (self-care) was a moderator of depression outcome among unguided users of *MoodGYM*.

Table 7.5. Potential treatment moderator variables considered in models of depression outcome. ✓= significant moderator; x = non-significant variable; shaded cells represent variables that were not investigated in the study.

		Current RCT (N=3,070)	Karyotaki meta-analysis (N=3,876)	Meyer RCT (N=396)
Demographic variables	Sex	x	x	x
	Age	x	x	x
	Educational level	x	x	x
	Marital status	x	x	
	Employment status	x	x	
	Daily Internet use	x		
Health-related variables	Previous use of CBT	x		
	Self-reported depression/anxiety	x		
	Previous mental health treatment	x		
	Smoking status	x		
	Previous use of Internet intervention	x		
	Level of physical activity	x		
	Mobility problems	x		
	Self-care problems	✓		
	Usual activities problems	x		
	Health state (0 to 100)	x		
	Current use of anti-depressants			✓
	Current psychotherapy			x
Clinical variables	Anxiety	x	x	
	Mental wellbeing (WEBWBS)	x		
	Baseline depression		x	x
	Age of depression onset			x
	MDD diagnosis			x

It is not clear why self-care was an important moderator of iCBT outcome whereas other dimensions related to everyday health status and quality of life (such as mobility problems or problems with usual activities) were not. It is, however, conceivable that individuals with self-care difficulties might experience particular impediments to successfully undertaking self-guided iCBT (e.g., regular computer use to access the program). Daily Internet use (or not) was tested separately and was not a

treatment modifier, although it did approach statistical significance (p = .08). However, Chi-square tests revealed no association between self-care problems and daily Internet use (χ^2 = 0.12, p = .73). The finding that self-care problems moderate the effect of unguided iCBT on depressive symptoms requires replication in future research, together with further exploration of the nature of the self-care problems. The confirmation of self-care status as a moderator of treatment effect might be important in providing advice to individuals about whether unguided iCBT is likely to be a useful treatment option. On the other hand, an understanding of the nature of this barrier might inform future iCBT interventions. It is possible, for example, that self-care status is a moderator of outcomes for unguided but not guided interventions, or that those with self-care problems might benefit from the presence of an assistant as they complete the intervention.

The current findings differed from those of the U.K.-based Button study of guided CBT [214] which found that those who were divorced, separated or widowed benefited the most from an iCBT intervention relative to the control. By contrast, the current study found that divorce was a predictor of outcomes for both the intervention and the control groups but was not a moderator of the treatment effect. This difference in findings may reflect the effect of personal input in the guided intervention. Alternatively, it may reflect the difference in the sample frame for the two studies with Button recruiting from primary care and the current study recruiting from community users of a health information website. Further moderator research in community and primary care settings and with both guided and unguided interventions is required.

7.4.4 Limitations

There are some limitations associated with the current study.

The study was exploratory, and the models constructed to test treatment effect modifiers did not control for other demographic or clinical characteristics. Further, no correction was made for the application of multiple statistical testing; it is, therefore, possible that the self-care moderator effect is spurious. Conversely, the sample size was powered to detect differences in outcome measures between intervention and control groups, and although large, may have been insufficient to detect moderator effects since the detection of an interaction effect requires a sample size approximately four times that required to detect a main effect [210, 211]. With an appropriately powered study, additional treatment modifiers might be identified, and a more complex model could be constructed to test potential moderators while controlling for other participant characteristics. It is also possible that the treatment effect was modified by a characteristic not measured in the current study. In

particular, concurrent anti-depressant medication use, which has been previously found to moderate unguided iCBT for depression, was not measured.

7.5 Conclusion

Participants in both the automated iCBT and control conditions were likely to have greater improvement in depressive symptoms if they were younger, had higher anxiety, had a lower level of mental wellbeing, had not previously accessed CBT or other mental health treatment, or were divorced. For those in the iCBT group only, depression outcomes were moderated by self-care status – i.e. whether the person had problems washing or dressing. Further research involving larger RCTs is required to verify this finding and to identify any other iCBT treatment modifiers and the impact of delivery modality (e.g., guided vs unguided). The findings of such studies have the potential to inform clinical practice guidelines for the treatment of depression in different groups of individuals and could be used as part of pre-intervention surveys which provide feedback about the suitability of programs such as *MoodGYM* for interested users and the conditions (e.g., guided or unguided) under which they may best be delivered. However, caution is required about extrapolating RCT results to the real-world context of spontaneous, community users.

8. Summary and conclusions

Depression is a highly prevalent disorder and is the leading cause of disability worldwide [1]. However, as previously noted, the majority of affected individuals do not access evidence-based treatment [3], due to a range of barriers including cost, lack of services, lack of knowledge, stigma and a preference for self-reliance [4]. The delivery of cognitive behaviour therapy over the Internet, if effective, has the potential to circumvent many of these barriers and provide an evidence-based resource to many who would otherwise not receive quality help. In particular, automated iCBT is highly scalable and can be provided at low marginal cost globally to large numbers of people – with the potential for a significant public health impact. In this context, this thesis has undertaken a detailed study of the *MoodGYM* program – an open-access, automated iCBT program for depression.

This chapter presents a summary of the results of the studies described in the thesis and then a discussion of the implications of these findings for public health. The chapter closes with potential future research directions.

8.1 Summary of results

8.1.1 Systematic reviews of iCBT for depression

The systematic reviews presented in this thesis provide evidence for the effectiveness of iCBT for depression. The first meta-review (Chapter 2) identified 46 existing systematic reviews of trials of iCBT for the treatment or prevention of depression, which were mostly assessed as being of high quality. Overall, the reviews indicated that iCBT was effective for the *treatment* of depression in adults and young people. Effect sizes from eight quantitative meta-analyses of iCBT treatment in adults ranged from small (*SMD* = 0.33) to large (*SMD* = 0.90), and a meta-analysis of treatment studies in young people reported a moderate overall effect size (*SMD* = 0.76). Qualitative reviews of iCBT for depression treatment in unrestricted populations (adults and youth) also reported positive results. There was less evidence for the use of iCBT for the *prevention* of depression. Only three meta-reviews focused exclusively on prevention: a review of iCBT for *indicated* prevention in adults found a moderate overall effect size (*SMD* = 0.46); a review of *universal* interventions in adults reported a small overall effect size (*SMD* = 0.35), although this included a trial of a non-iCBT intervention; and one review found evidence for prevention of depression in young people but did not distinguish between different types of prevention. In addition, three reviews of mixed trials found evidence for the use of iCBT with women in the perinatal period.

Where reported, participant adherence was generally lower in treatment trials of unguided iCBT compared to guided iCBT. However, several recent reviews found no difference in the relative effectiveness of unguided and guided iCBT – although these comparisons were made from pooled results, with the included trials varying on a range of factors. One review did consider head-to-head trials of guided vs unguided iCBT and found a non-significant difference in favour of guided iCBT. Importantly, for the context of the current thesis, there was evidence that unguided iCBT is effective for the treatment of depression, with a recent individual participant data analysis yielding a small effect size (*SMD* = 0.27) based on data from 13 trials of unguided iCBT [28].

Only a small amount of research has been conducted on the cost-effectiveness of iCBT for depression. Three relevant reviews were identified, collectively including only four studies. The evidence from these studies provides preliminary evidence that both guided and unguided iCBT are cost-effective compared to health intervention thresholds.

Having established that iCBT is generally effective for reducing depressive symptoms, a systematic qualitative review of unguided, automated iCBT for depression was undertaken (Chapter 3). In total, 17 RCTs were identified, with most assessed as having a low risk of bias. The majority of studies (64.7%) reported positive results, with effect sizes for these trials ranging from 0.16 to 0.76. There was good evidence for the effectiveness of automated iCBT within community settings: of the nine RCTs conducted with participants from the community, eight demonstrated effectiveness. The one community-based trial which did not report positive results involved a randomly selected community sample who were screened for depression, whereas the other trials were self-selected individuals who responded to advertisements about the study. A trial of automated iCBT with crisis helpline callers and another with university students also yielded positive results, suggesting that this may be an effective intervention approach for these populations with unmet need. There were mixed results for the four trials conducted in primary care. Finally, a trial with secondary school students showed that unguided iCBT was not effective, suggesting that the use of iCBT in school settings requires some form of human involvement (for example, supervision by a teacher).

8.1.2 Real-world users: Observational studies of automated iCBT for depression

Given the positive results for effectiveness trials of automated iCBT for depression among self-selected individuals from the community, the studies in Chapters 4 and 5 examined real-world users of the open-access *MoodGYM* program. These studies examined a large dataset of users (*N* = 317,354) who registered with the program between 2009 and 2014 and indicated that they were looking for help for themselves. As summarised in Table 8.1 (a) below, users were mostly

female, highly educated and had previously accessed other evidence-based treatments for depression. Nearly half logged in from the United Kingdom, and a large proportion was referred to the program by a health professional. Overall, users had high baseline depression and anxiety symptoms compared to the general population.

Comparisons between users from the top four countries (United Kingdom, Australia, Canada and the United States) revealed that the highest proportions of male users and users who found the program via a web-link or online search were from the U.S., whereas the percentage of users who were referred by a health professional was highest among those from the U.K. There was also a significantly higher percentage of users from rural areas among those from the U.K. or Australia compared to the U.S. or Canada.

The study presented in Chapter 4 also investigated whether user characteristics could predict referral by a health professional or higher baseline depression. Users who were male, aged 40-49 years old or from the United Kingdom, Canada or rural areas were most likely to be referred by a health professional [Table 8.1b (i)]. Higher baseline depression scores were predicted by self-reporting a history of depression, having previously tried more depression treatments, having a lower level of education, being referred by a health professional, and logging in from the United Kingdom [Table 8.1b (ii)].

Table 8.1. Summary of user characteristics and predictors among community MoodGYM users.

(a) User characteristics (N=417,354).

User characteristic	**%**	**User characteristic**	**%**
Female	68.1%	Motivated to work through the program	24.7%
Modal age: 20-39 years	31.3%	Not used CBT but want to try it	52.1%
Country:		Previously accessed:	
United Kingdom	49.5%	≥ 1 evidence-based depression treatment	88.3%
Australia	20.8%	≥ 3 evidence-based depression treatments	57.6%
Canada	6.5%	Anti-depressant medication	50.6%
Rural location	24.6%	**User characteristic**	***M, SD***
University degree	40.5%	Baseline Goldberg Depression (*n*=372,432)	6.29, 1.88
Health professional referral	42.7%	Baseline Goldberg Anxiety (*n*=386,981)	6.42, 1.98

(b) Predictors of (i) health professional referral (n=340,313); (ii) baseline depression (n=301,373); and (iii) depression change (n=106,882).

(i) Health professional referral	**(ii) Higher baseline depression**	**(iii) Greater depression improvement**
• Male • Aged 40-49 years • From the U.K. or Canada • Rural • Lower education level	• Self-reported history of depression • More treatments previously tried • Lower education level • Health professional referral • From the U.K.	• Higher baseline depression • Lower baseline anxiety • More modules completed • More time from pre- to post-test • Lower baseline dysfunctional thinking • Self-reported history of depression • Motivated to work through program • Higher education level • Female • Not used anti-depressants • From the U.S. • Not being aged ≤ 19 years • Rural

(c) Predictors of adherence: (i) commencing Module 1 (n=340,313); and (ii) completing at least 1 vs no modules (n=278,151).

(i) Commencing Module 1	**(ii) Completing 1 or ≥ 2 vs 0 modules**
• Younger age • Not used CBT but want to try it • Motivated to work through program • Self-reported history of depression • Rural location • From the U.K. • Health professional referral • Female • Higher education level	• Motivated to work through program • Younger age • Not used CBT but want to try it • Health professional referral • From the U.K. or Australia[a] • Education level[b] • No self-reported history of depression[c] • Female[d] • Higher baseline depression[d] • Higher baseline anxiety

[a] *U.K. users were more likely to complete ≥ 2 vs 0 modules than users from Australia.*

[b] *Lower education predicted 1 vs 0 modules, but higher education predicted ≥ 2 vs 0 modules.*

[c] *Only a predictor of completing ≥ 2 vs 0 modules.*

[d] *Only a predictor of completing 1 vs 0 module.*

User adherence to the program was considered in Chapter 5. Of those who registered, 81.4% completed the initial assessment surveys and continued to commence Module 1. A logistic regression showed that users who were younger, who had not used CBT but wanted to try it, who were motivated to complete the program, or who had a self-reported history of depression were more likely to commence Module 1 [see Table 8.1c (i) above]. Of those users who commenced Module 1, 44% did not complete this module, 39% completed only the first module, 17% completed two or more modules, and a small number completed the entire program (2.9%). A multinomial regression analysis was conducted to investigate which registration or baseline clinical variables predicted completion of either one, or two or more modules compared to zero modules. The most important predictor of module completion was whether users indicated at registration that they were motivated to complete the program, with motivated users 61% more likely to complete two or more modules compared to no modules. The results of this analysis are summarised in Table 8.1c (ii). On average, users logged in 2.54 times (*SD* = 3.13) and spent 79.6 minutes (*SD* = 108.28) on the program module webpages.

For those users who completed at least two depression symptom assessments (*n* = 130,426; 38.4% of those users who commenced Module 1; 31.3% of the total sample), there was a significant, small decrease in mean depression symptom scores from initial assessment to the last assessment undertaken [mean score change = 0.58 (95% CI: 0.57, 0.59); Cohen's *d* = 0.28]. At an individual level this magnitude of change is not likely to represent a meaningful or clinical benefit, although such an effect could translate into a major benefit at a population level. A hierarchical multivariate regression analysis was undertaken to explore whether improvements in depression symptoms were predicted by initial clinical assessment scores, registration characteristics or adherence variables, although caution is required in drawing conclusions from this analysis since only the sub-group of participants who completed at least two assessments were included. The results of the final model (including all types of independent variables) indicated that baseline clinical scores and adherence to the intervention were the strongest predictors of depression improvement, with those with higher initial depression and those who completed more modules showing the largest positive improvements [Table 8.1b (iii)].

8.1.3 Choice and intervention variations: Partially Randomised Preference Trial

The Partially Randomised Preference Trial (*N* = 23,988) described in Chapter 6 investigated which *MoodGYM* modules are required to achieve a reduction in depression symptoms in community users. In addition, the study explored the role of user preference in outcomes and adherence. Personal community users who agreed to participate in the study were asked whether they would like to

choose a program variant or if they would agree to be randomised. A logistic regression analysis found that participants who were older, more educated, more anxious or less depressed were more likely to choose their program variant (see Table 8.2).

Table 8.2. Summary of predictors of choice status (Chapter 6, n=19,316).

Chose condition
• Higher education level • Older age • Higher baseline anxiety • Lower baseline depression • From North America • Previous use of anti-depressant medication • No self-reported history of depression

Completion of the assigned or chosen program was not predicted by choice status. However, participants who chose their condition were significantly more likely to complete at least two depression tests, and choice status remained a significant predictor after controlling for the condition received. Choosing a condition rather than being randomised was also a significant predictor of larger positive change in depression symptoms; however, choice status no longer significantly predicted depression improvement after controlling for condition.

The relative effectiveness of six different combinations of the existing MoodGYM modules was examined in the RCT arm of the trial (n = 7,794), although conclusions drawn from the results of this study are limited by the small number of participants who completed all of the modules in the condition (14.7%) and the potential confounding effect of the differing length of the conditions. Linear mixed effects model analysis for repeated measures revealed that the two most effective conditions were Condition 4 (Modules 1, 2 and 5) and Condition 5 (Modules 1, 2, 3 and 5); however, differences between conditions were small. Further contrasts found significantly greater improvements in those variations which included Module 2 compared to those that did not. In particular, Condition 4 (Modules 1, 2 and 5) was significantly superior to Condition 3 (Modules 1, 4 and 5) and Condition 2 (Module 1 and 5). The results suggest that the 'extended CBT' modules, Module 1 and 2, may be the most important MoodGYM components for depression outcomes.

8.1.4 Moderators of automated iCBT for depression

Treatment moderators of unguided *MoodGYM* for depression were examined in Chapter 7, in an RCT of community participants located in England who were recruited from a government health information website (*N* = 3,070). Although the observational study reported in Chapter 5 explored

predictors of depression outcome in spontaneous users, a comparative control condition is required to determine if there are characteristics which moderate outcome for iCBT users only (but not for the sample more generally). A series of separate linear mixed effect models were conducted to explore whether any of the baseline demographic, clinical or health-related variables moderated improvement in depression outcome for those in the CBT condition. This exploratory analysis found that the only variable which was associated with depression outcome in the *MoodGYM* condition but not the control condition was self-reported ability to wash or dress (i.e., no problems with self-care). A multivariate linear mixed model analysis also found that a number of variables predicted depression improvement across both groups; however, several of these were inconsistent with those identified in Chapter 5 – for example, younger participants and those with higher anxiety improved more in this RCT whereas older and less anxious users improved more in the observational study.

8.2 Implications of research findings

The studies in this thesis provide encouraging evidence for the potential of open-access, automated iCBT for depression to impact public health. To the best of our knowledge, the review of automated iCBT (Chapter 3) is the first to focus specifically on trials of iCBT for depression which do not include any direct human input during the intervention period. The results suggest that self-selected community users may be the most appropriate target group for such programs, which is consistent with their open-access delivery. The observational studies of real-world *MoodGYM* users are the largest implementation studies for depression iCBT to date and provide the following important insights into the uptake of an open-access automated intervention.

First, automated iCBT appears to be a viable as a low intensity stepped care resource to which health professionals can refer clients, and the adoption of automated iCBT into health systems may be a critical driver of uptake. Nearly half of the users were from the United Kingdom, and a large proportion of British users were referred by a health professional. This likely reflects the integration of CCBT into the U.K. health system as part of clinical guidelines for the management of depression, and the associated promotion through the National Health Service system. Users were generally highly educated and mostly female; however, males and those with a lower education level were more likely to be referred by a health professional. This demonstrates the importance of the awareness of such resources by clinicians since it appears some sub-groups of the population may be more likely to access automated interventions on the advice of their doctor. Interestingly, a study which investigated the use of a GP testimonial integrated into the *MoodGYM* registration process found that it did not affect the rate of user registrations [225], suggesting that an in-person recommendation may be necessary. This in turn requires the systematic implementation of programs

designed to increase practitioner awareness of evidence-based programs. In Australia, the 'e-Mental Health in Practice' (eMHPrac) initiative[8] was launched in 2014 and provides training and support for health practitioners about e-mental health (including the use of *MoodGYM* and similar programs). Although it commenced after the current study, the eMHPrac project will evaluate health referral numbers and whether, as expected, it is associated with an increase in the practitioner referral rate to *MoodGYM* in Australia. Potentially, such programs provide a template for delivery in other countries.

Second, an open-access automated intervention can provide community-based treatment for those who cannot or would prefer not to access other forms of help. The studies suggest that *MoodGYM* attracts users with an unmet need since the spontaneous users had high levels of baseline depression symptoms and the program was accessed by a high percentage of users from rural or remote regions (e.g., 21.1% of Australian users) where other treatment options are typically not readily available.

Third, these programs are of interest to individuals of all ages. Approximately one-third of the spontaneous users were aged 40 years or older, despite the fact that *MoodGYM* was originally designed for young people.

Fourth, automated iCBT may be less suitable for users with lower levels of education. Overall open-access *MoodGYM* attracted users who were highly educated. This may be because more highly educated individuals have increased exposure to information about depression, and hence a higher level of depression literacy, including how to seek help. However, spontaneous users with lower educational attainment were also less likely to progress past the initial assessment surveys and exhibited less improvement in depression symptoms. The program is relatively 'text heavy', and as such, is likely to be less accessible for those with lower levels of reading literacy. Given that less educated users were also more likely to have higher levels of depression symptoms, there is a need to develop interventions which rely less on text, perhaps through the use of video or graphical communication. Content could also be adapted to reduce the reading grade level and to teach simplified CBT techniques.

Finally, it is clear that low adherence among spontaneous users is an issue of substantial importance. By their nature, open-access programs are likely to demonstrate lower adherence because they facilitate access by those who have no intention of or no commitment to completing the program but who rather visit in order to 'have a look'. Nevertheless, there is a clear need to consider strategies for

[8] Information about eMHPrac can be found at http://www.emhprac.org.au.

increasing user engagement, particularly given that completion of more modules most strongly predicted depression outcomes. The predictors identified in the observational study provide some potential strategies. For example, given that they predicted depression outcomes after controlling for adherence, measures to *motivate* and *educate* interested users about iCBT and what to expect from *MoodGYM* may boost the program's impact. In Australia, the Commonwealth Department of Health's 'Head to Health' website[9], launched in 2017, is designed to assist individuals to locate suitable e-mental resources for their situation (including *MoodGYM*); however, currently there is no explanation or primer about different types of interventions such as iCBT. This kind of portal (a trusted government platform for individuals seeking help) may be an appropriate setting to implement pre-intervention measures such as motivational interviewing [176] or the provision of brief interactive information about what to expect from CBT, which could increase subsequent user engagement and improve outcomes for a larger public health impact.

Changes to the intervention itself may also increase user engagement and require further exploration. For example, reducing the amount of text through the use of video and graphics may increase engagement across users of all educational levels, as might the use of automated, scheduled and possibly tailored text messages or emails designed to support users and invite them to re-engage with specific content. Content which anticipates barriers to practicing changes and staying involved in the program could also be added – for example using character stories to model coping strategies. In addition, delivery of the program using a responsive web design so that the program is accessible on devices with different sized screens (including smartphones and tablets) could facilitate ongoing user access.

The PRPT results suggest that implementing user choice may prove another means to increase user engagement with multi-module open-access programs. The RCT arm of this trial also found that the first two *MoodGYM* modules are the most critical program components for depression outcomes. Hence, to increase engagement and symptom improvement, the *MoodGYM* program itself might be altered so that users can more freely choose which modules and sections they complete after the core content (Modules 1 and 2). Tailored recommendations could also be incorporated, since users who are more highly depressed, younger or less educated may be less willing or able to choose between alternative options. Simultaneously, pre-intervention measures could emphasise the importance of the first two modules, with users encouraged to complete these at a minimum in order to achieve a positive outcome.

[9] Head to Health is available at https://headtohealth.gov.au.

The exploratory moderator analyses presented in Chapter 7 suggest that users who have difficulty caring for themselves may not benefit from automated iCBT for depression. It could be that such users require human guidance in order to benefit from iCBT. However, this result requires validation through future research. Only a small number of RCTs of guided or unguided iCBT for depression have employed moderator (as opposed to predictor) analyses to investigate who benefits most from such programs. Such approaches need to be more commonly included in planning effectiveness studies. An understanding of for whom iCBT is most effective could inform clinical practice guidelines for the use of iCBT programs which involve different levels of guidance and assist health practitioners to decide which individuals might benefit from referral to iCBT. Information about for whom automated iCBT is the most benefit could also inform pre-intervention materials or automated recommendation algorithms.

8.3 Conclusions and future research directions

This thesis has shown that automated iCBT can be effective for reducing depression symptoms in community users and that it can be successfully delivered as an open-access public health resource. Nearly half of the spontaneous users of the *MoodGYM* program (2009-2014) were from the United Kingdom, and a large proportion of these users were referred by a health professional, suggesting that automated iCBT can successfully contribute to an existing health system. However, further research is required to investigate the drivers of referral. As discussed above, more research into moderators of treatment effect for iCBT is also needed, to inform clinical practice guidelines and to assist health professionals to make decisions about the appropriateness of referring individual patients to automated iCBT. Additional studies are also required to explore attitudes of potential users, which types of users engage with automated iCBT services, and how best to promote open-access programs to different sub-groups of users.

Maximising user engagement with automated iCBT is also a critical domain of ongoing and future investigation. As previously noted, *MoodGYM* users were likely to complete more of the program if they were motivated to do so at the time of registration. Hence, there is a need to explore ways to motivate users before they use the intervention, as well as techniques to keep users engaged as they progress through the program. Including user choice as part of the intervention experience may increase engagement; however, further research with spontaneous users is required.

Future work must also focus on how the effectiveness of automated programs for depression can be maximised. There is evidence that the first two *MoodGYM* modules are the critical components for depression symptom improvement. However, newer intervention optimisation approaches could be

employed, such as a MOST framework with SMART design, to investigate which and how modules should be delivered beyond the core modules, for different types of users. Along with analyses of predictors of depression outcome, this could then inform algorithms which make real-time recommendations depending on the characteristics of the user, how they have engaged with the program, or their response to the intervention (i.e. whether their symptoms have reduced).

Comparative studies with other open-access iCBT depression programs are also required to determine whether the user characteristics and predictors of referral, adherence and depression improvement in spontaneous users of *MoodGYM* are specific to the intervention or generalisable to automated iCBT.

The public health impact of *MoodGYM* could also be further examined within the RE-AIM Evaluation model, which employs five dimensions (Reach, Efficacy, Adoption, Implementation and Maintenance) to calculate a public health impact score of the population-based effect of the intervention [226, 227]. Future implementation science research could also extend the current findings – for example by exploring how to most effectively promote the availability and appropriate use of iCBT depression programs to health professionals, or how to best market the program to prospective users while taking account of localised differences in healthcare systems. Further research is also required to explore the potential for the program to be culturally adapted for use in non-Western contexts, building on encouraging preliminary evidence that a Chinese (Mandarin) language version of *MoodGYM* may be effective for treating Chinese patients with depression [228].

Finally, although there is evidence that the open-access global delivery of *MoodGYM* is a viable public health intervention which has achieved a meaningful impact, its cost-effectiveness needs to be quantified through further research. Clearly, the intervention is highly scalable and has reached large numbers of individuals with elevated depression symptoms. However, its impact has been magnified in countries where iCBT has been integrated into existing health systems and promoted through health services. Evidence for the cost-effectiveness could help to demonstrate to policy makers the value of automated iCBT as a public health intervention and promote its further integration into health systems worldwide.

9. References

1. World Health Organization, *Depression and other common mental disorders: global health estimates*. 2017.

2. Vigo, D.M.D., G.P. Thornicroft, and R.P. Atun, *Estimating the true global burden of mental illness.* Lancet Psychiatry, 2016. **3**(2): p. 171-178.

3. Olfson, M., C. Blanco, and S.C. Marcus, *Treatment of Adult Depression in the United StatesTreatment of Adult Depression in the United StatesTreatment of Adult Depression in the United States.* JAMA Internal Medicine, 2016. **176**(10): p. 1482-1491.

4. Prins, M., et al., *Perceived need for mental health care and barriers to care in the Netherlands and Australia.* Social Psychiatry and Psychiatric Epidemiology, 2011. **46**(10): p. 1033-1044.

5. Cuijpers, P., et al., *How effective are cognitive behavior therapies for major depression and anxiety disorders? A meta-analytic update of the evidence.* World Psychiatry, 2016. **15**(3): p. 245-258.

6. Grist, R. and K. Cavanagh, *Computerised Cognitive Behavioural Therapy for Common Mental Health Disorders, What Works, for Whom Under What Circumstances? A Systematic Review and Meta-analysis.* Journal of Contemporary Psychotherapy, 2013. **43**(4): p. 243-251.

7. Griffiths, K., L. Farrer, and H. Christensen, *Clickety-click: E-Mental Health Train on Track.* Australasian Psychiatry, 2007. **15**(2): p. 100-108.

8. Andrews, G., et al., *Computer Therapy for the Anxiety and Depressive Disorders Is Effective, Acceptable and Practical Health Care: A Meta-Analysis.* PLoS One, 2010. **5**(10): p. e13196.

9. Nelson, E.-L. and A.B. Duncan, *Cognitive Behavioral Therapy Using Televideo.* Cognitive and Behavioral Practice, 2015. **22**(3): p. 269-280.

10. Cavanagh, K. and D.A. Shapiro, *Computer treatment for common mental health problems.* Journal of Clinical Psychology, 2004. **60**(3): p. 239-251.

11. Griffiths, K.M. and H. Christensen, *Internet-based mental health programs: A powerful tool in the rural medical kit.* Australian Journal of Rural Health, 2007. **15**(2): p. 81-87.

12. Munoz, R.F., et al., *Massive Open Online Interventions: A Novel Model for Delivering Behavioral-Health Services Worldwide.* Clinical Psychology Science, 2015: p. 1-12.

13. Christensen, H., K.M. Griffiths, and C. Groves, *MoodGYM Training Program: Clinician's Manual*. 2004, Canberra: Centre for Mental Health Research.

14. Gulliver, A., et al., *Internet-Based Interventions to Promote Mental Health Help-Seeking in Elite Athletes: An Exploratory Randomized Controlled Trial.* Journal of Medical Internet Research, 2012. **14**(3): p. e69.

15. Xu, Z., et al., *Effectiveness of interventions to promote help-seeking for mental health problems: systematic review and meta-analysis.* Psychological Medicine, 2018. **48**(16): p. 2658-2667.

16. Clarke, G., et al., *Overcoming depression on the Internet (ODIN): a randomized controlled trial of an Internet depression skills intervention program.* Journal of Medical Internet Research, 2002. **4**(3): p. E14.

17. Griffiths, K.M. and H. Christensen, *Review of randomised controlled trials of internet interventions for mental disorders and related conditions.* Clinical Psychologist, 2006. **10**(1): p. 16-29.

18. Kaltenthaler, E., et al., *Computerized cognitive behavior theraphy for depression and anxiety update: A systematic review and economic evaluation.* Health Technology Assessment, 2006. **10**.

19. Foroushani, P.S., J. Schneider, and N. Assareh, *Meta-review of the effectiveness of computerised CBT in treating depression.* BMC Psychiatry, 2011. **11**(1): p. 131.

20. Rogers, M.A., et al., *Internet-Delivered Health Interventions That Work: Systematic Review of Meta-Analyses and Evaluation of Website Availability.* Journal of Medical Internet Research, 2017. **19**(3): p. e90.

21. Liberati, A., et al., *The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration.* PLOS Medicine, 2009. **6**(7): p. e1000100.

22. Shea, B.J., et al., *Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews.* BMC Medical Research Methodology, 2007. **7**(1): p. 10.

23. Kim, G., Y.Z. Cho, and S.K. Baik, *Assessment for Risk of Bias in Systematic Reviews and Meta-Analyses in the Field of Hepatology.* Gut and Liver, 2015. **9**(6): p. 701-706.

24. Pieper, D., et al., *Systematic review found AMSTAR, but not R(evised)-AMSTAR, to have good measurement properties.* Journal of Clinical Epidemiology, 2015. **68**(5): p. 574-83.

25. Higgins, J.P., et al., *The Cochrane Collaboration's tool for assessing risk of bias in randomised trials.* British Medical Journal, 2011. **343**: p. d5928.

26. Sztein, D.M., et al., *Efficacy of cognitive behavioural therapy delivered over the Internet for depressive symptoms: A systematic review and meta-analysis.* Journal of Telemedicine and Telecare, 2017: p. 1357633x17717402.

27. Lau, Y., et al., *Therapist-Supported Internet-Based Cognitive Behavior Therapy for Stress, Anxiety, and Depressive Symptoms Among Postpartum Women: A Systematic Review and Meta-Analysis.* Journal of Medical Internet Research, 2017. **19**(4): p. e138.

28. Karyotaki, E., et al., *Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: A meta-analysis of individual participant data.* JAMA Psychiatry, 2017. **74**(4): p. 351-359.

29. Josephine, K., et al., *Internet- and mobile-based depression interventions for people with diagnosed depression: A systematic review and meta-analysis.* Journal of Affective Disorders, 2017. **223**: p. 28-40.

30. Zhou, T., et al., *Internet-based cognitive behavioural therapy for subthreshold depression: a systematic review and meta-analysis.* BMC Psychiatry, 2016. **16**: p. 356.

31. Sander, L., L. Rausch, and H. Baumeister, *Effectiveness of Internet-Based Interventions for the Prevention of Mental Disorders: A Systematic Review and Meta-Analysis.* JMIR Mental Health, 2016. **3**(3): p. e38.

32. Lee, E.W., et al., *Web-based interventions for prevention and treatment of perinatal mood disorders: a systematic review.* BMC Pregnancy and Childbirth, 2016. **16**: p. 38.

33. Ebert, D.D., et al., *Does Internet-based guided-self-help for depression cause harm? An individual participant data meta-analysis on deterioration rates and its moderators in randomized controlled trials.* Psychological Medicine, 2016. **46**(13): p. 2679-93.

34. Brown, M., et al., *Gamification and Adherence to Web-Based Mental Health Interventions: A Systematic Review.* JMIR Mental Health, 2016. **3**(3): p. e39.

35. de Bitencourt Machado, D., et al., *A psychodynamic perspective on a systematic review of online psychotherapy for adults.* British Journal of Psychotherapy, 2016. **32**(1): p. 79-108.

36. Ashford, M.T., E.K. Olander, and S. Ayers, *Computer- or web-based interventions for perinatal mental health: A systematic review.* Journal of Affective Disorders, 2016. **197**: p. 134-46.

37. Karyotaki, E., et al., *Predictors of treatment dropout in self-guided web-based interventions for depression: An 'individual patient data' meta-analysis.* Psychological Medicine, 2015. **45**(13): p. 2717-2726.

38. Ebert, D.D., et al., *Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials.* PLoS One, 2015. **10**(3): p. e0119895.

39. Donker, T., et al., *Economic evaluations of Internet interventions for mental health: a systematic review.* Psychological Medicine, 2015. **45**(16): p. 3357-76.

40. Clarke, A.M., T. Kuosmanen, and M.M. Barry, *A systematic review of online youth mental health promotion and prevention interventions.* Journal of Youth and Adolescence, 2015. **44**(1): p. 90-113.

41. Ye, X., et al., *Effectiveness of internet-based interventions for children, youth, and young adults with anxiety and/or depression: a systematic review and meta-analysis.* BMC Health Services Research, 2014. **14**: p. 313.

42. van Ballegooijen, W., et al., *Adherence to Internet-based and face-to-face cognitive behavioural therapy for depression: a meta-analysis.* PLoS One, 2014. **9**(7): p. e100674.

43. Saddichha, S., et al., *Online interventions for depression and anxiety - a systematic review.* Health Psychology and Behavioral Medicine, 2014. **2**(1): p. 841-881.

44. Rice, S.M., et al., *Online and social networking interventions for the treatment of depression in young people: a systematic review.* Journal of Medical Internet Research, 2014. **16**(9): p. e206.

45. Reyes-Portillo, J.A., et al., *Web-based interventions for youth internalizing problems: a systematic review.* Journal of the American Academy of Child and Adolescent Psychiatry, 2014. **53**(12): p. 1254-1270.e5.

46. Davies, E.B., R. Morriss, and C. Glazebrook, *Computer-delivered and web-based interventions to improve depression, anxiety, and psychological well-being of university students: a systematic review and meta-analysis.* Journal of Medical Internet Research, 2014. **16**(5): p. e130.

47. Baumeister, H., et al., *The impact of guidance on Internet-based mental health interventions — A systematic review.* Internet Interventions, 2014. **1**(4): p. 205-215.

48. Arnberg, F.K., et al., *Internet-delivered psychological treatments for mood and anxiety disorders: a systematic review of their efficacy, safety, and cost-effectiveness.* PLoS One, 2014. **9**(5): p. e98118.

49. Andersson, G., et al., *Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis.* World Psychiatry, 2014. **13**(3): p. 288-95.

50. Paul, C., et al. *The impact of web-based approaches on psychosocial health in chronic physical and mental health conditions*. Health Education Research, 2013. **28**, 450-71.

51. Cowpertwait, L. and D. Clarke, *Effectiveness of Web-based Psychological Interventions for Depression: A Meta-analysis.* International Journal of Mental Health and Addiction, 2013. **11**(2): p. 247-268.

52. Richards, D. and T. Richardson, *Computer-based psychological treatments for depression: A systematic review and meta-analysis.* Clinical Psychology Review, 2012. **32**(4): p. 329-342.

53. Muresan, V., G.H. Montgomery, and D. David, *Emotional outcomes and mechanisms of change in online cognitive-behavioral interventions: A quantitative meta-analysis of clinical controlled studies.* Journal of Technology in Human Services, 2012. **30**(1): p. 1-13.

54. Hedman, E., B. Ljotsson, and N. Lindefors, *Cognitive behavior therapy via the Internet: a systematic review of applications, clinical efficacy and cost-effectiveness.* Expert Review of Pharmacoeconomics & Outcomes Research, 2012. **12**(6): p. 745-64.

55. Titov, N., *Internet-delivered psychotherapy for depression in adults.* Current Opinion in Psychiatry, 2011. **24**(1): p. 18-23.

56. Wade, A.G., *Use of the internet to assist in the treatment of depression and anxiety: a systematic review.* Primary Care Companion to the Journal of Clinical Psychiatry, 2010. **12**(4).

57. Griffiths, K.M., L. Farrer, and H. Christensen, *The efficacy of internet interventions for depression and anxiety disorders: a review of randomised controlled trials.* Medical Journal of Australia, 2010. **192**(11 Suppl): p. S4-11.

58. García-Lizana, F. and I. Muñoz-Mayorga, *Telemedicine for Depression: A Systematic Review.* Perspectives in Psychiatric Care, 2010. **46**(2): p. 119-126.

59. Calear, A.L. and H. Christensen, *Review of internet-based prevention and treatment programs for anxiety and depression in children and adolescents.* Medical Journal of Australia, 2010. **192**(11 Suppl): p. S12-4.

60. Christensen, H., K.M. Griffiths, and L. Farrer, *Adherence in Internet intervention for anxiety and depression: Systematic review.* Journal of Medical Internet Research, 2009. **11**(2): p. 1-16.

61. Andersson, G. and P. Cuijpers, *Internet-based and other computerized psychological treatments for adult depression: a meta-analysis.* Cognitive Behaviour Therapy, 2009. **38**(4): p. 196-205.

62. Amstadter, A.B., et al., *Internet-based interventions for traumatic stress-related mental health problems: a review and suggestion for future research.* Clinical Psychology Review, 2009. **29**(5): p. 410-20.

63. Kaltenthaler, E., et al., *Computerised cognitive-behavioural therapy for depression: Systematic review.* British Journal of Psychiatry, 2008. **193**.

64. Barak, A., et al., *A Comprehensive Review and a Meta-Analysis of the Effectiveness of Internet-Based Psychotherapeutic Interventions.* Journal of Technology in Human Services, 2008. **26**(2-4): p. 109-160.

65. Spek, V., et al., *Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis.* Psychological Medicine, 2007. **37**(3): p. 319-28.

66. Gellatly, J., et al., *What makes self-help interventions effective in the management of depressive symptoms? Meta-analysis and meta-regression.* Psychological Medicine, 2007. **37**(9): p. 1217-1228.

67. Mrazek, P.J. and R.J. Haggerty, *Reducing risks for mental disorders: frontiers for preventive intervention research*. 1994, Washington, DC: National Academy Press.

68. Donkin, L., et al., *A Systematic Review of the Impact of Adherence on the Effectiveness of e-Therapies.* Journal of Medical Internet Research, 2011. **13**(3): p. e52.

69. Saha, S., et al., *The art and science of incorporating cost effectiveness into evidence-based recommendations for clinical preventive services.* American Journal of Preventive Medicine, 2001. **20**(3, Supplement 1): p. 36-43.

70. Ramsey, S., et al., *Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task Force report.* Value Health, 2005. **8**(5): p. 521-33.

71. Owens, D.K., *Interpretation of Cost-Effectiveness Analyses.* Journal of General Internal Medicine, 1998. **13**(10): p. 716-717.

72. Brown, M.M., et al., *Health Care Economic Analyses and Value-Based Medicine.* Survey of Ophthalmology, 2003. **48**(2): p. 204-223.

73. McCabe, C., K. Claxton, and A. Culyer, *The NICE Cost-Effectiveness Threshold: What it is and What that Means*. Vol. 26. 2008. 733-44.

74. Hollinghurst, S., et al., *Cost-effectiveness of therapist-delivered online cognitive-behavioural therapy for depression: randomised controlled trial.* British Journal of Psychiatry, 2010. **197**(4): p. 297-304.

75. Warmerdam, L., et al., *Cost-utility and cost-effectiveness of internet-based treatment for adults with depressive symptoms: randomized trial.* Journal of Medical Internet Research, 2010. **12**(5): p. e53.

76. Gerhards, S.A., et al., *Economic evaluation of online computerised cognitive-behavioural therapy without support for depression in primary care: randomised trial.* British Journal of Psychiatry, 2010. **196**(4): p. 310-8.

77. Phillips, R., et al., *Randomized controlled trial of computerized cognitive behavioural therapy for depressive symptoms: effectiveness and costs of a workplace intervention.* Psychological Medicine, 2014. **44**(4): p. 741-752.

78. Christensen, H., K.M. Griffiths, and A.F. Jorm, *Delivering interventions for depression by using the internet: randomised controlled trial.* British Medical Journal, 2004. **328**(7434): p. 265.

79. Kessler, R.C., et al., *Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication.* Archives of General Psychiatry, 2005. **62**(6): p. 593-602.

80. Kennard, B.D., et al., *Relapse and recurrence in pediatric depression.* Child and Adolescent Psychiatric Clinics of North America, 2006. **15**(4): p. 1057-79, xi.

81. Keller, M.B., et al., *Long-term outcome of episodes of major depression. Clinical and public health significance.* JAMA, 1984. **252**(6): p. 788-92.

82. Rozental, A., G. Andersson, and P. Carlbring, *In the Absence of Effects: An Individual Patient Data Meta-Analysis of Non-response and Its Predictors in Internet-Based Cognitive Behavior Therapy.* Frontiers in Psychology, 2019. **10**: p. 589-589.

83. Rozental, A., et al., *For better or worse: An individual patient data meta-analysis of deterioration among participants receiving Internet-based cognitive behavior therapy.* Journal of Consulting and Clinical Psychology, 2017. **85**(2): p. 160-177.

84. Christensen, H. and A. Mackinnon, *The Law of Attrition Revisited.* Journal of Medical Internet Research, 2006. **8**(3): p. e20.

85. Kessler, R.C. and E.J. Bromet, *The epidemiology of depression across cultures.* Annual Review of Public Health, 2013. **34**: p. 119-138.

86. Vos, T., et al., *Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010.* The Lancet, 2012. **380**(9859): p. 2163-96.

87. Christensen, H., K.M. Griffiths, and A.F. Jorm, *Delivering interventions for depression by using the internet: randomised controlled trial.* BMJ, 2004. **328**(7434): p. 265.

88. Bown, M.J. and A.J. Sutton, *Quality Control in Systematic Reviews and Meta-analyses.* European Journal of Vascular and Endovascular Surgery, 2010. **40**(5): p. 669-677.

89. Garg, A.X., D. Hackam, and M. Tonelli, *Systematic Review and Meta-analysis: When One Study Is Just not Enough.* Clinical Journal of the American Society of Nephrology, 2008. **3**(1): p. 253.

90. Eldridge, S.M., et al., *Defining Feasibility and Pilot Studies in Preparation for Randomised Controlled Trials: Development of a Conceptual Framework.* PLoS One, 2016. **11**(3): p. e0150205.

91. Cuijpers, P., A. van Straten, and L. Warmerdam, *Problem solving therapies for depression: a meta-analysis.* European Psychiatry, 2007. **22**(1): p. 9-15.

92. Mira, A., et al., *An Internet-based program for depressive symptoms using human and automated support: a randomized controlled trial.* Neuropsychiatr Dis Treat, 2017. **13**: p. 987-1006.

93. Noguchi, R., et al., *Effects of five-minute internet-based cognitive behavioral therapy and simplified emotion-focused mindfulness on depressive symptoms: a randomized controlled trial.* BMC Psychiatry, 2017. **17**(1): p. 85.

94. Lokman, S., et al., *Complaint-Directed Mini-Interventions for Depressive Complaints: A Randomized Controlled Trial of Unguided Web-Based Self-Help Interventions.* Journal of Medical Internet Research, 2017. **19**(1): p. e4.

95. Silverstone, P.H., et al., *Depression outcomes in adults attending family practice were not improved by screening, stepped-care, or online CBT during a 12-week study when compared to controls in a randomized trial.* Frontiers in Psychiatry, 2017. **8**.

96. Romero-Sanchiz, P., et al., *Economic evaluation of a guided and unguided internet-based CBT intervention for major depression: Results from a multi-center, three-armed randomized controlled trial conducted in primary care.* PLoS One, 2017. **12**(2): p. e0172741.

97. Montero-Marin, J., et al., *An Internet-Based Intervention for Depression in Primary Care in Spain: A Randomized Controlled Trial.* Journal of Medical Internet Research, 2016. **18**(8): p. e231.

98. Meyer, B., et al., *Effects of an Internet intervention (Deprexis) on severe depression symptoms: Randomized controlled trial.* Internet Interventions, 2015. **2**(1): p. 48-59.

99. Twomey, C., et al., *A randomized controlled trial of the computerized CBT programme, MoodGYM, for public mental health service users waiting for interventions.* British Journal of Clinical Psychology, 2014. **53**(4): p. 433-50.

100. Lillevoll, K.R., et al., *Uptake and adherence of a self-directed internet-based mental health intervention with tailored e-mail reminders in senior high schools in Norway.* BMC Psychiatry, 2014. **14**: p. 14.

101. Mohr, D.C., et al., *A Randomized Controlled Trial Evaluating a Manualized TeleCoaching Protocol for Improving Adherence to a Web-Based Intervention for the Treatment of Depression.* PLoS One, 2013. **8**(8): p. e70086.

102. Lintvedt, O.K., et al., *Evaluating the effectiveness and efficacy of unguided internet-based self-help intervention for the prevention of depression: a randomized controlled trial.* Clinical Psychology & Psychotherapy, 2013. **20**(1): p. 10-27.

103. Powell, J., et al., *Effectiveness of a web-based cognitive-behavioral tool to improve mental well-being in the general population: randomized controlled trial.* Journal of Medical Internet Research, 2012. **15**(1): p. e2.

104. Berger, T., et al., *Internet-Based Treatment of Depression: A Randomized Controlled Trial Comparing Guided with Unguided Self-Help.* Cognitive Behaviour Therapy, 2011. **40**(4): p. 251-266.

105. Farrer, L., et al., *Internet-Based CBT for Depression with and without Telephone Tracking in a National Helpline: Randomised Controlled Trial.* PLoS One, 2011. **6**(11): p. e28099.

106. Meyer, B., et al., *Effectiveness of a Novel Integrative Online Treatment for Depression (Deprexis): Randomized Controlled Trial.* Journal of Medical Internet Research, 2009. **11**(2): p. e15.

107. de Graaf, L.E., et al., *Clinical effectiveness of online computerised cognitive–behavioural therapy without support for depression in primary care: randomised trial.* British Journal of Psychiatry, 2009. **195**(1): p. 73-80.

108. Spek, V., et al., *Internet-based cognitive behavioural therapy for subthreshold depression in people over 50 years old: a randomized controlled clinical trial.* Psychological Medicine, 2007. **37**(12): p. 1797-1806.

109. Spek, V., et al., *One-year follow-up results of a randomized controlled clinical trial on internet-based cognitive behavioural therapy for subthreshold depression in people over 50 years.* Psychological Medicine, 2008. **38**(5): p. 635-9.

110. Sheehan, D.V., et al., *The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10.* Journal of Clinical Psychiatry, 1998. **59 Suppl 20**: p. 22-33;quiz 34-57.

111. Andrews, G. and T. Slade, *Interpreting scores on the Kessler Psychological Distress Scale (K10).* Australian & New Zealand Journal of Public Health, 2001. **25**(6): p. 494-497.

112. Klein, J.P., et al., *Time to remission from mild to moderate depressive symptoms: One year results from the EVIDENT-study, an RCT of an internet intervention for depression.* Behaviour Research and Therapy, 2017. **97**: p. 154-162.

113. Klein, J.P., et al., *Effects of a Psychological Internet Intervention in the Treatment of Mild to Moderate Depressive Symptoms: Results of the EVIDENT Study, a Randomized Controlled Trial.* Psychotherapy and Psychosomatics, 2016. **85**(4): p. 218-228.

114. Uher, R. and B. Pavlova, *Long-term effects of depression treatment.* Lancet Psychiatry, 2016. **3**(2): p. 95-96.

115. Ibrahim, A.K., et al., *A systematic review of studies of depression prevalence in university students.* Journal of Psychiatric Research, 2013. **47**(3): p. 391-400.

116. Givens, J.L. and J. Tjia, *Depressed medical students' use of mental health services and barriers to use.* Academic Medicine, 2002. **77**(9): p. 918-21.

117. Calear, A.L., et al., *The YouthMood Project: a cluster randomized controlled trial of an online cognitive behavioral program with adolescents.* Journal of Consulting and Clinical Psychology, 2009. **77**(6): p. 1021-32.

118. Christensen, H., et al., *Online randomized controlled trial of brief and full cognitive behaviour therapy for depression.* Psychological Medicine, 2006. **36**(12): p. 1737-1746.

119. Gerhards, S.A.H., et al., *Improving adherence and effectiveness of computerised cognitive behavioural therapy without support for depression: A qualitative study on patient experiences.* Journal of Affective Disorders, 2011. **129**(1): p. 117-125.

120. Jo, B., *Statistical power in randomized intervention studies with noncompliance.* Psychological Methods, 2002. **7**(2): p. 178-193.

121. Lachin, J.M., *Fallacies of Last Observation Carried Forward.* Clinical Trials, 2016. **13**(2): p. 161-168.

122. Blankers, M., J.M.W. Koeter, and M.G. Schippers, *Missing Data Approaches in eHealth Research: Simulation Study and a Tutorial for Nonmathematically Inclined Researchers.* Journal of Medical Internet Research, 2010. **12**(5): p. e54.

123. Muñoz, F.R., *Using Evidence-Based Internet Interventions to Reduce Health Disparities Worldwide.* Journal of Medical Internet Research, 2010. **12**(5): p. e60.

124. Muñoz, R.F., *Using Evidence-Based Internet Interventions to Reduce Health Disparities Worldwide.* Journal of Medical Internet Research, 2010. **12**(5): p. e60.

125. Curran, G.M., et al., *Effectiveness-implementation Hybrid Designs: Combining Elements of Clinical Effectiveness and Implementation Research to Enhance Public Health Impact.* Medical Care, 2012. **50**(3): p. 217-226.

126. Karyotaki, E., H. Riper, and P. Cuijpers, *Internet-based self-help interventions for depression in routine care—reply.* JAMA Psychiatry, 2017. **74**(8): p. 853-853.

127. Fleming, T., et al., *Beyond the Trial: Systematic Review of Real-World Uptake and Engagement With Digital Self-Help Interventions for Depression, Low Mood, or Anxiety.* Journal of Medical Internet Research, 2018. **20**(6): p. e199.

128. Lattie, E.G., et al., *Uptake and usage of IntelliCare: A publicly available suite of mental health and well-being apps.* Internet Interventions, 2016. **4**: p. 152-158.

129. Carpenter, J., et al., *Seeing the "Big" Picture: Big Data Methods for Exploring Relationships Between Usage, Language, and Outcome in Internet Intervention Data.* Journal of Medical Internet Research, 2016. **18**(8): p. e241.

130. Christensen, H., K.M. Griffiths, and A. Korten, *Web-based cognitive behavior therapy: analysis of site usage and changes in depression and anxiety scores.* Journal of Medical Internet Research, 2002. **4**(1): p. e3.

131. Christensen, H., et al., *A Comparison of Changes in Anxiety and Depression Symptoms of Spontaneous Users and Trial Participants of a Cognitive Behavior Therapy Website.* Journal of Medical Internet Research, 2004. **6**(4): p. e46.

132. Christensen, H., et al., *Free range users and one hit wonders: community users of an Internet-based cognitive behaviour therapy program.* Australian & New Zealand Journal of Psychiatry, 2006. **40**(1): p. 59-62.

133. Batterham, P.J., et al., *Predictors of adherence among community users of a cognitive behavior therapy website.* Patient Preference and Adherence, 2008. **2**: p. 97-105.

134. Neil, A.L., et al., *Predictors of adherence by adolescents to a cognitive behavior therapy website in school and community-based settings.* Journal of Medical Internet Research, 2009. **11**(1): p. e6.

135. Lara, M.A., et al., *A four-year experience with a Web-based self-help intervention for depressive symptoms in Mexico.* Pan American Journal of Public Health, 2014. **35**(5-6): p. 399-406.

136. Goldberg, D., et al., *Detecting anxiety and depression in general medical settings.* British Medical Journal, 1988. **297**(6653): p. 897-899.

137. Parslow, R.A., et al., *The warpy thoughts scale: a new 20-item instrument to measure dysfunctional attitudes.* Cognitive Behaviour Therapy, 2006. **35**(2): p. 106-16.

138. National Institute for Clinical Excellence (NICE), *Guidance on the use of computerised cognitive behavioural therpay for anxiety and depression*, in *NICE Technology Appraisal Guidance No. 51*. 2002.

139. Christensen, H., K.M. Griffiths, and K. Evans, *e-Mental Health in Australia: Implications of the Internet and Related Technologies for Policy*. 2002, Commonwealth Department of Health and Ageing: Canberra.

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

141. Lumley, T., et al., *The importance of the normality assumption in large public health data sets.* Annual Review of Public Health, 2002. **23**(1): p. 151-169.

142. Paul, P., M.L. Pennell, and S. Lemeshow, *Standardizing the power of the Hosmer-Lemeshow goodness of fit test in large data sets.* Statistics in Medicine, 2013. **32**(1): p. 67-80.

143. Christensen, H. and K.M. Griffiths, *MoodGYM: Overcoming depression with CBT and other effective therapies*. 2011, London: Vermilion.

144. Albert, P.R., *Why is depression more prevalent in women?* Journal of Psychiatry & Neuroscience, 2015. **40**(4): p. 219-221.

145. Patten, S.B., et al., *Descriptive epidemiology of major depression in Canada.* Canadian Journal of Psychiatry, 2006. **51**(2): p. 84-90.

146. Statista. *Percentage of population using the internet in Mexico from 2000 to 2017*. [accessed 7/11/2018]; Available from: https://www.statista.com/statistics/209112/number-of-internet-users-per-100-inhabitants-in-mexico-since-2000/.

147. eMarketer. *Mexico's Growing Internet Audience Buoyed by Young Users*. [accessed 7/11/2018]; Available from: https://www.emarketer.com/Article/Mexicos-Growing-Internet-Audience-Buoyed-by-Young-Users/1011632.

148. Magaard, J.L., et al., *Factors associated with help-seeking behaviour among individuals with major depression: A systematic review.* PLoS One, 2017. **12**(5): p. e0176730.

149. World Bank. *Rural population (% of total population)*. [accessed 10 September 2018]; Available from: https://data.worldbank.org/indicator/SP.RUR.TOTL.ZS.

150. Driessen, E. and S.D. Hollon, *Cognitive Behavioral Therapy for Mood Disorders: Efficacy, Moderators and Mediators.* The Psychiatric Clinics of North America, 2010. **33**(3): p. 537-555.

151. Pybis, J., et al., *The comparative effectiveness and efficiency of cognitive behaviour therapy and generic counselling in the treatment of depression: evidence from the 2nd UK National Audit of psychological therapies.* BMC Psychiatry, 2017. **17**(1): p. 215.

152. Zhang, Z., et al., *The effect of CBT and its modifications for relapse prevention in major depressive disorder: a systematic review and meta-analysis.* BMC Psychiatry, 2018. **18**(1): p. 50.

153. Australian Bureau of Statistics, *National Survey of Mental Health and Wellbeing: Summary of Results, 2007*. Cat. no. (4326.0). 2008, Canberra: ABS.

154. Kessler, R.C., et al., *The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R).* JAMA, 2003. **289**(23): p. 3095-105.

155. McManus, S., et al., *Mental health and wellbeing in England: Adult Psychiatric Morbidity Survey 2014*. 2016, Leeds: NHS Digital.

156. Rotter, J.B., *Social learning and clinical psychology*. 1954, New York: Prentice-Hall.

157. Shifrer, D. and A. Sutton, *Region-Urbanicity Differences in Locus of Control: Social Disadvantage, Structure, or Cultural Exceptionalism?* Sociological Inquiry, 2014. **84**(4): p. 570-600.

158. Jackson, H., et al., *Mental health problems in rural contexts: what are the barriers to seeking help from professional providers.* Australian Psychologist, 2007. **42**(2).

159. Anthony, J.S., et al., *Conditions That Influence a Primary Care Clinician's Decision to Refer Patients for Depression Care.* Rehabilitation Nursing, 2010. **35**(3): p. 113-22.

160. CARMHA Faculty of Health Sciences, Simon Fraser University, *Family Physician Guide for depression, anxiety disorders, early psychosis, and substance use disorders*. 2008, Ministry of Health, Mental Health and Addictions Branch: British Columbia.

161. Call, J.B. and K. Shafer, *Gendered Manifestations of Depression and Help Seeking Among Men.* American Journal of Men's Health, 2018. **12**(1): p. 41-51.

162. Griffiths, K.M., H. Christensen, and A.F. Jorm, *Predictors of depression stigma.* BMC Psychiatry, 2008. **8**(1): p. 25-25.

163. De Silva, T., et al., *General practitioners' experiences and perceptions of mild moderate depression management and factors influencing effective service delivery in rural Australian communities: a qualitative study.* International Journal of Mental Health Systems, 2017. **11**(1): p. 54.

164. Lindner, P., et al., *Who seeks ICBT for depression and how do they get there? Effects of recruitment source on patient demographics and clinical characteristics.* Internet Interventions, 2015. **2**(2): p. 221-225.

165. Jacobi, F., et al., *Mental disorders in the general population : Study on the health of adults in Germany and the additional module mental health (DEGS1-MH).* Nervenarzt, 2014. **85**(1): p. 77-87.

166. Fox, S. and M. Duggan, *Mobile Health 2012*. 2012, Pew Internet & American Life Project: Washington, D.C.

167. Lin, M., H.C. Lucas, and G. Shmueli, *Research Commentary—Too Big to Fail: Large Samples and the p-Value Problem.* Information Systems Research, 2013. **24**(4): p. 906-917.

168. Ebert, D. and H. Baumeister, *Internet-based self-help interventions for depression in routine care.* JAMA Psychiatry, 2017.

169. Kelders, S.M., et al., *Persuasive System Design Does Matter: A Systematic Review of Adherence to Web-Based Interventions.* Journal of Medical Internet Research, 2012. **14**(6): p. e152.

170. Waller, R. and S. Gilbody, *Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence.* Psychological Medicine, 2009. **39**(05): p. 705-712.

171. Donkin, L. and N. Glozier, *Motivators and Motivations to Persist With Online Psychological Interventions: A Qualitative Study of Treatment Completers.* Journal of Medical Internet Research, 2012. **14**(3): p. e91.

172. Wilhelmsen, M., et al., *Motivation to persist with internet-based cognitive behavioural treatment using blended care: a qualitative study.* BMC Psychiatry, 2013. **13**(1): p. 296.

173. El Alaoui, S., et al., *Predicting Outcome in Internet-Based Cognitive Behaviour Therapy for Major Depression: A Large Cohort Study of Adult Patients in Routine Psychiatric Care.* PLoS One, 2016. **11**(9): p. e0161191.

174. Vatcheva, K.P., et al., *Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies.* Epidemiology, 2016. **6**(2): p. 227.

175. Ritterband, L.M., et al., *A Behavior Change Model for Internet Interventions.* Annals of Behavioral Medicine, 2009. **38**(1): p. 18-27.

176. Rollnick, S., et al., *Motivational interviewing.* British Medical Journal, 2010. **340**: p. c1900.

177. Andersson, G. and N. Titov, *Advantages and limitations of Internet-based interventions for common mental disorders.* World Psychiatry, 2014. **13**(1): p. 4-11.

178. Lara, M.A. and M. Tiburcio Sainz, *Dropout in self-guided web-based interventions for depression can be predicted by several variables.* Evidence-based Mental Health, 2016. **19**(1).

179. Cuijpers, P., *The challenges of improving treatments for depression.* JAMA, 2018.

180. Christensen, H., P. Batterham, and A. Calear, *Online interventions for anxiety disorders.* Current Opinion in Psychiatry, 2014. **27**(1): p. 7-13.

181. Twomey, C. and G. O'Reilly, *Effectiveness of a freely available computerised cognitive behavioural therapy programme (MoodGYM) for depression: Meta-analysis.* Australian & New Zealand Journal of Psychiatry, 2016.

182. Høifødt, R.S., et al., *Predictors of Response to Web-Based Cognitive Behavioral Therapy With High-Intensity Face-to-Face Therapist Guidance for Depression: A Bayesian Analysis.* Journal of Medical Internet Research, 2015. **17**(9): p. e197-e197.

183. Hamilton, K.E. and K.S. Dobson, *Cognitive therapy of depression: Pretreatment patient predictors of outcome.* Clinical Psychology Review, 2002. **22**(6): p. 875-893.

184. Jorm, A.F., *Evaluations of Web-Based Psychological Interventions Need to Avoid the Re-Test Artefact.* Australian & New Zealand Journal of Psychiatry, 2009. **43**(9): p. 877-877.

185. Corrigan, P.W. and M.S. Salzer, *The conflict between random assignment and treatment preference: implications for internal validity.* Evaluation and Program Planning, 2003. **26**(2): p. 109-121.

186. Howard, L. and G. Thornicroft, *Patient preference randomised controlled trials in mental health research.* British Journal of Psychiatry, 2006. **188**(APR.): p. 303-304.

187. Kowalski, C.J. and A.J. Mrdjenovich, *Patient Preference Clinical Trials: why and when they will sometimes be preferred.* Perspectives in Biology and Medicine, 2013. **56**(1): p. 18-35.

188. Brewin, C.R. and C. Bradley, *Patient preferences and randomised clinical trials.* British Medical Journal, 1989. **299**(6694): p. 313-315.

189. King, M., et al., *Impact of participant and physician intervention preferences on randomized trials: A systematic review.* JAMA, 2005. **293**(9): p. 1089-1099.

190. Laugharne, R. and P. Stefan, *Trust, choice and power in mental health.* Social Psychiatry and Psychiatric Epidemiology, 2006. **41**(11): p. 843-52.

191. Swift, J.K. and J.L. Callahan, *The impact of client treatment preferences on outcome: a meta-analysis.* Journal of Clinical Psychology, 2009. **65**(4): p. 368-381.

192. Lindhiem, O., et al., *Client Preferences Affect Treatment Satisfaction, Completion, and Clinical Outcome: A Meta-Analysis.* Clinical Psychology Review, 2014. **34**(6): p. 506-517.

193. Muñoz, F.R., et al., *From Online Randomized Controlled Trials to Participant Preference Studies: Morphing the San Francisco Stop Smoking Site into a Worldwide Smoking Cessation Resource.* Journal of Medical Internet Research, 2012. **14**(3): p. e64.

194. Vandelanotte, C., et al., *Do Participants? Preferences for Mode of Delivery (Text, Video, or Both) Influence the Effectiveness of a Web-Based Physical Activity Intervention?* Journal of Medical Internet Research, 2012. **14**(1): p. e37.

195. Andersson, G., et al., *Can the Patient Decide Which Modules to Endorse? An Open Trial of Tailored Internet Treatment of Anxiety Disorders.* Cognitive Behaviour Therapy, 2011. **40**(1): p. 57-64.

196. Johansson, R., et al., *Choosing between Internet-based psychodynamic versus cognitive behavioral therapy for depression: a pilot preference study.* BMC Psychiatry, 2013. **13**(1): p. 268.

197. Gemmell, I. and G. Dunn, *The statistical pitfalls of the partially randomized preference design in non-blinded trials of psychological interventions.* International Journal of Methods in Psychiatric Research, 2011. **20**(1): p. 1-9.

198. Schafer, J.L., J.L. Schafer, and J.W. Graham, *Missing data: Our view of the state of the art.* Psychological Methods, 2002. **7**(2): p. 147-177.

199. Hindmarch, T., M. Hotopf, and G.S. Owen, *Depression and decision-making capacity for treatment or research: a systematic review.* BMC Medical Ethics, 2013. **14**(1): p. 54.

200. Morán-Sánchez, I., A. Maurandi-López, and M.D. Pérez-Cárceles, *Assessment of Motivations and Willingness to Participate in Research of Outpatients With Anxiety, Mood, and Psychotic Disorders.* Journal of Empirical Research on Human Research Ethics, 2018. **13**(5): p. 546-560.

201. Dunn, L.B., et al., *Worth the Risk? Relationship of Incentives to Risk and Benefit Perceptions and Willingness to Participate in Schizophrenia Research.* Schizophrenia Bulletin, 2009. **35**(4): p. 730-737.

202. Karlawish, J.H.T., D.J. Casarett, and B.D. James, *Alzheimer's Disease Patients' and Caregivers' Capacity, Competency, and Reasons to Enroll in an Early-Phase Alzheimer's Disease Clinical Trial.* Journal of the American Geriatrics Society, 2002. **50**(12): p. 2019-2024.

203. Collins, L.M., K.C. Kugler, and M.V. Gwadz, *Optimization of Multicomponent Behavioral and Biobehavioral Interventions for the Prevention and Treatment of HIV/AIDS.* AIDS and Behavior, 2016. **20 Suppl 1**(0 1): p. S197-S214.

204. Collins, L.M., I. Nahum-Shani, and D. Almirall, *Optimization of behavioral dynamic treatment regimens based on the sequential, multiple assignment, randomized trial (SMART).* Clinical Trials, 2014. **11**(4): p. 426-434.

205. Torgerson, D. and J.K. Moffett, *Patient Preference and Validity of Randomized Controlled Trials.* JAMA, 2005. **294**(1): p. 41-42.

206. Piper, M.E., et al., *Identifying effective intervention components for smoking cessation: a factorial screening experiment.* Addiction, 2016. **111**(1): p. 129-141.

207. Kamper, S.J., et al., *Treatment-based subgroups of low back pain: A guide to appraisal of research studies and a summary of current evidence.* Best Practice & Research Clinical Rheumatology, 2010. **24**(2): p. 181-191.

208. Kraemer, H.C., *Discovering, comparing, and combining moderators of treatment on outcome after randomized clinical trials: a parametric approach.* Statistics in Medicine, 2013. **32**(11): p. 1964-1973.

209. Kraemer, H.C., et al., *Mediators and Moderators of Treatment Effects in Randomized Clinical Trials.* Archives of General Psychiatry, 2002. **59**(10): p. 877-883.

210. Brookes, S.T., et al., *Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test.* Journal of Clinical Epidemiology, 2004. **57**(3): p. 229-36.

211. Leon, A.C. and M. Heo, *Sample Sizes Required to Detect Interactions between Two Binary Fixed-Effects in a Mixed-Effects Linear Regression Model.* Computational Statistics & Data Analysis, 2009. **53**(3): p. 603-608.

212. MacKinnon, D.P., *Integrating Mediators and Moderators in Research Design.* Research on Social Work Practice, 2011. **21**(6): p. 675-681.

213. Donker, T., et al., *Predictors and moderators of response to internet-delivered Interpersonal Psychotherapy and Cognitive Behavior Therapy for depression.* Journal of Affective Disorders, 2013. **151**(1): p. 343-51.

214. Button, K.S., et al., *Factors associated with differential response to online cognitive behavioural therapy.* Social Psychiatry and Psychiatric Epidemiology, 2012. **47**(5): p. 827-33.

215. Warmerdam, L., et al., *Predicting outcome of Internet-based treatment for depressive symptoms.* Psychotherapy Research, 2013. **23**(5): p. 559-567.

216. Spek, V., et al., *Predictors of outcome of group and internet-based cognitive behavior therapy.* Journal of Affective Disorders, 2007. **105**(1): p. 137-145.

217. Radloff, L.S., *The CES-D Scale: A Self-Report Depression Scale for Research in the General Population.* Applied Psychological Measurement, 1977. **1**(3): p. 385-401.

218. Donker, T., et al., *Brief self-rated screening for depression on the Internet.* Journal of Affective Disorders, 2010. **122**(3): p. 253-259.

219. Zich, J.M., C.C. Attkisson, and T.K. Greenfield, *Screening for depression in primary care clinics: the CES-D and the BDI.* International Journal of Psychiatry in Medicine, 1990. **20**(3): p. 259-77.

220. Spitzer, R.L., et al., *A brief measure for assessing generalized anxiety disorder: the GAD-7.* Archives of Internal Medicine, 2006. **166**(10): p. 1092-7.

221. Tennant, R., et al., *The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation.* Health and Quality of Life Outcomes, 2007. **5**: p. 63.

222. Rabin, R. and F. de Charro, *EQ-5D: a measure of health status from the EuroQol Group.* Annals of Medicine, 2001. **33**(5): p. 337-43.

223. Milton, K., F.C. Bull, and A. Bauman, *Reliability and validity testing of a single-item physical activity measure.* British Journal of Sports Medicine, 2011. **45**(3): p. 203-8.

224. Office for National Statistics. *Census 2011*. [accessed 20 December 2018]; Available from: https://webarchive.nationalarchives.gov.uk/20160105165938/http://www.ons.gov.uk/ons/guide-method/census/2011/index.html.

225. Healey, B.J., K.M. Griffiths, and K. Bennett, *The effect of programme testimonials on registrations for an online cognitive behaviour therapy intervention: a randomised trial.* Digital Health, 2017. **3**.

226. Vis, C., et al., *Improving Implementation of eMental Health for Mood Disorders in Routine Practice: Systematic Review of Barriers and Facilitating Factors.* JMIR Mental Health, 2018. **5**(1): p. e20-e20.

227. Glasgow, R.E., T.M. Vogt, and S.M. Boles, *Evaluating the public health impact of health promotion interventions: the RE-AIM framework.* American Journal of Public Health, 1999. **89**(9): p. 1322-1327.

228. Yeung, A., et al., *Outcomes of an online computerized cognitive behavioral treatment program for treating chinese patients with depression: A pilot study.* Asian Journal of Psychiatry, 2018. **38**: p. 102-107.

Appendix A: Adapted AMSTAR Checklist Items

Adapted AMSTAR Checklist Items

Item 1 Was the research question (i.e., research purpose) clearly specified?

Item 2 Was there duplicate study selection and data extraction?
There should be at least two independent data extractors and a consensus procedure for disagreements should be in place.

Item 3 Was a comprehensive literature search performed?
The report must include years and databases used [e.g. Central, EMBASE, and MEDLINE]. Key words and/or MeSH terms must be stated and where feasible the search strategy should be provided.

Item 4 Were inclusion/exclusion criteria reported clearly?

Item 5 Was a list of studies (included and excluded) provided?
To be assessed 'Yes' on this item, at a minimum a list of included studies should be provided.

Item 6 Were the characteristics of the included studies provided?
In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes.

Item 7 Was the scientific quality of the included studies assessed and documented?
'A priori' methods of assessment should be provided.

Item 8 Was the scientific quality of the included studies used appropriately in formulating conclusions?
The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations.

Item 9 Were the methods used to combine the findings of studies appropriate?
For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e. Chi-squared test for homogeneity). If heterogeneity exists a random effects model should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e. is it sensible to combine?).

Item 10 Was the likelihood of publication bias assessed?
An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test).

Item 11	Was the conflict of interest stated? *Potential sources of support should be clearly acknowledged in both the systematic review and the included studies. To be assessed 'Yes' on this item, at a minimum conflict of interest and sources of support should be listed for the systematic review.*

Following the method applied by Kim et al. to assess the methodological quality of systematic reviews in the field of hepatology[1], two of the original AMSTAR items[2] were modified for the current purposes. The original AMSTAR Item 1, "Was an 'a priori' design provided?", required reference to a protocol, ethics approval or pre-determined published research objectives. Since as noted by Kim et al. most published systematic reviews do not have pre-published protocols, this item was changed to, "Was the research question (i.e., research purpose) clearly specified?". In addition, Item 4, "Was the status of publication (i.e. grey literature) used as an inclusion criterion" was modified to "Were inclusion/exclusion criteria reported clearly?", Kim et al. arguing that it is "difficult to extract data from grey or from unpublished literature sources".

[1] Kim, G., Y.Z. Cho, and S.K. Baik, *Assessment for Risk of Bias in Systematic Reviews and Meta-Analyses in the Field of Hepatology.* Gut and Liver, 2015. **9**(6): p. 701-706.
[2] Shea, B.J., et al., *Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews.* BMC Medical Research Methodology, 2007. **7**(1): p. 10.

Appendix B: Additional analyses

Results from additional regression analyses for Chapter 5 are described below.

B.1. Predicting completion of at least one depression test

A logistic regression was conducted to investigate which registration variables independently predicted completion of at least one depression test, for those users who registered from one of the top four countries (*n* = 340,313). All independent variables were entered simultaneously. The model test statistics are displayed at Table B.1 and the odds ratios for the independent variables within the regression model are shown at Table B.2.

Table B.1. Logistic regression model test statistics.

	Chi-square (*df*, *p*)	**-2 Log likelihood**	**Cox & Snell R^2**	**Nagelkerke R^2**
Model (gender, age, country, location, education, referral, previous depression, motivated, experience of CBT, treatment history)	2,271.65 (23, <.001)	228,657.08	.007	.014

Table B.2. Logistic regression results for predicting completion of at least one depression test (n=340,313).

	Odds Ratio	95% CI	*p*	Wald
Gender				
Female	**1.11*****	[1.08, 1.13]	<.001	71.25
Age group ** reference group: ≥ 50 years*				
≤ 19 years	**1.69*****	[1.61, 1.78]	<.001	446.39
20 – 29 years	**1.54*****	[1.49, 1.60]	<.001	612.15
30 – 39 years	**1.34*****	[1.29, 1.38]	<.001	282.57
40 – 49 years	**1.21*****	[1.17, 1.25]	<.001	110.74
Country ** reference group: Australia*				
United Kingdom	**1.10*****	[1.07, 1.13]	<.001	48.40
United States	**0.88*****	[0.84, 0.93]	<.001	25.78
Canada	**0.92*****	[0.88, 0.96]	<.001	14.44
Location				
Rural/remote	**1.12*****	[1.09, 1.15]	<.001	75.82
Education level ** reference group: Higher degree*				
Some secondary	**0.80*****	[0.76, 0.85]	<.001	67.43
Secondary only	**0.94****	[0.89, 0.98]	.008	7.07
Trade /certificate	**0.90*****	[0.86, 0.93]	<.001	31.43
Undergraduate	0.99	[0.94, 1.05]	.791	0.17
Diploma	**0.88*****	[0.84, 0.92]	<.001	33.59
Bachelor's degree	**0.96***	[0.92, 0.99]	.017	5.72
Health professional referral				
Referred	**1.05*****	[1.03, 1.08]	<.001	20.33
Depression history				
Previous depression	**1.10*****	[1.06, 1.13]	<.001	31.55
Motivation level				
Motivated	**1.18*****	[1.15, 1.22]	<.001	151.54
Experience of CBT ** reference group: Not sure what CBT is*				
Tried, useful	1.00	[0.96, 1.05]	.832	0.04
Tried, not useful	**1.09****	[1.03, 1.15]	.002	9.90
Not tried, want to	**1.23*****	[1.19, 1.26]	<.001	202.56
Not tried, don't want to	0.98	[0.92, 1.04]	.515	0.42
Treatment history				
Number of treatments	**0.97*****	[0.96, 0.97]	<.001	118.08

$^{**}p < .01$, $^{***}p < .001$

B.2. Predicting number of modules completed

The multinomial logistic regression for predicting completion of one module compared to zero modules, and two or modules compared to zero modules (section 5.3.1, pp. 184-188), was rerun replacing the 20-item Warpy Thoughts score with the total summed score of the 42 item Warpy Thoughts Quiz. The model test statistics are shown in Table B.3 and the odds ratios for the independent variables in the regression model are displayed in Table B.4.

Table B.3. Multinomial logistic regression model test statistics.

	Chi-square (*df*, *p*)	**-2 Log likelihood**	**Cox & Snell R^2**	**Nagelkerke R^2**
Model (gender, age, country, education, referral, previous depression, motivated, experience of CBT, baseline depression, baseline anxiety, baseline 42 item Warpy Thoughts, number of treatments)	4,641.69 (50, <.001)	559,411.78	.017	.019

Table B.4. Multinomial regression results for predicting completion of one module compared to zero modules, and two or modules compared to zero modules (n=278,151).

	One module vs zero modules				2-5 modules vs. zero modules			
	Odds Ratio	**95% CI**	***p***	**Wald**	**Odds Ratio**	**95% CI**	***p***	**Wald**
Gender								
Female	**1.03****	[1.01, 1.05]	.005	8.05	1.01	[0.98, 1.03]	.538	0.38
Age group* *reference group: ≥ 50 years*								
≤ 19 years	**1.45*****	[1.39, 1.50]	<.001	365.05	**1.45*****	[1.38, 1.52]	<.001	222.51
20 – 29 years	**1.25*****	[1.21, 1.28]	<.001	236.48	**1.08*****	[1.04, 1.11]	<.001	15.73
30 – 39 years	**1.10*****	[1.07, 1.13]	<.001	44.37	**0.95****	[0.92, 0.99]	.008	7.09
40 – 49 years	1.02	[0.99, 1.05]	.200	1.64	**0.96***	[0.92, 1.00]	.029	4.76
Country * *reference group: Australia*								
United Kingdom	1.00	[0.98, 1.02]	.790	0.07	**1.03***	[1.00, 1.06]	.024	5.08
United States	**0.86*****	[0.83, 0.89]	<.001	58.96	**0.90*****	[0.85, 0.95]	<.001	16.96
Canada	**0.90*****	[0.87, 0.93]	<.001	37.67	**0.93****	[0.89, 0.97]	.001	10.31
Education level * *reference group: Higher degree*								
Some secondary	**1.09*****	[1.05, 1.14]	<.001	17.92	**0.88*****	[0.83, 0.93]	<.001	22.42
Secondary only	**1.08*****	[1.05, 1.12]	<.001	19.63	**0.81*****	[0.77, 0.85]	<.001	81.76
Trade /certificate	**1.07*****	[1.04, 1.10]	<.001	18.86	**0.80*****	[0.77, 0.83]	<.001	141.90
Undergraduate	**1.09*****	[1.05, 1.13]	<.001	17.91	**0.85*****	[0.80, 0.89]	<.001	43.23
Diploma	1.02	[0.98, 1.05]	.291	1.11	**0.80*****	[0.77, 0.84]	<.001	102.86
Bachelor's degree	1.02	[0.99, 1.05]	.124	2.36	**0.96***	[0.92, 0.99]	.011	6.47

Table B.4. Multinomial regression results for predicting completion of one module compared to zero modules, and two or modules compared to zero modules (n=278,151) (continued).

	One module vs zero modules				2-5 modules vs. zero modules			
	Odds Ratio	**95% CI**	***p***	**Wald**	**Odds Ratio**	**95% CI**	***p***	**Wald**
Health professional referral								
Referred	**1.11*****	[1.09, 1.13]	<.001	144.73	**1.21*****	[1.18, 1.23]	<.001	261.98
History of depression								
Previous depression	1.01	[0.98, 1.03]	.613	0.26	**0.89*****	[0.86, 0.92]	<.001	51.83
Motivation level								
Motivated	**1.16*****	[1.14, 1.18]	<.001	223.03	**1.61*****	[1.57, 1.65]	<.001	1496.19
Experience of CBT ** reference group: Not sure what it is*								
Tried, useful	**0.94*****	[0.91, 0.97]	<.001	14.24	**1.18*****	[1.13, 1.23]	<.001	55.66
Tried, not useful	**0.95***	[0.91, 0.99]	.012	6.30	**1.11*****	[1.05, 1.17]	<.001	12.99
Not tried, want to	**1.05*****	[1.03, 1.07]	<.001	18.66	**1.30*****	[1.26, 1.34]	<.001	310.44
Not tried, don't want to	**0.95***	[0.91, 1.00]	.040	4.21	1.00	[0.94, 1.07]	.967	0.00
Goldberg Depression (baseline)	**1.02*****	[1.02, 1.03]	<.001	61.94	1.00	[0.99, 1.01]	.780	0.08
Goldberg Anxiety (baseline)	**1.01*****	[1.01, 1.02]	<.001	17.73	1.01	[1.00, 1.01]	.073	3.21
Warpy Thoughts (baseline)	**1.0005****	[1.0001, 1.0008]	.007	7.24	**1.0006****	[1.0001, 1.0010]	.009	6.73
No. of previous treatments	**0.98*****	[0.98, 0.99]	<.001	49.45	1.00	[0.99, 1.01]	.727	0.12

*p < .05, **p < .01, ***p < .001

B.3. Predicting depression symptom changes

The hierarchical multivariate regression model for predicting depression symptom changes in users (section 5.3.2, pp. 197-200) was re-run, replacing the 20-item higher factor Warpy Thoughts Score with the three dysfunctional thinking factors (relationships factor, entitlements factor, achievements factor). All three factors were entered simultaneously in the first block, along with baseline depression and anxiety scores. As per the previous model, demographic and other registration characteristics were entered in the second block, and process or adherence variables were added in the final block. The model test statistics are displayed in Table B.5 and the standardised Beta coefficients for the independent variables are shown in Table B.6.

Table B.5. Multivariate regression model test statistics for depression change score

	R^2	R^2 change	*F* change	Sig *F* change
Step 1 (baseline depression, anxiety, dysfunctional thinking factors)	0.089	0.089	2,085.28	<.001
Step 2 (Step 1 + gender, age, education, country, location, motivated, depression history, experience of CBT, no. treatments, anti-depressants)	0.104	0.016	80.85	<.001
Step 3 (Step 2 + no. modules + time between tests)	0.144	0.040	1,644.40	<.001

Table B.6. Hierarchical multivariate regression predicting depression change score (n=106,882).

	Step 1		Step 2		Step 3	
	β	*P*	β	*p*	β	*p*
Depression (baseline)	**0.319*****	<.001	**0.345*****	<.001	**0.354*****	<.001
Anxiety (baseline)	**-0.147*****	<.001	**-0.140*****	<.001	**-0.135*****	<.001
WTS Relationships factor (baseline)	**-0.072*****	<.001	**-0.067*****	<.001	**-0.074*****	<.001
WTS Entitlements factor (baseline)	**-0.030*****	<.001	**-0.023*****	<.001	**-0.014*****	<.001
WTS Achievements factor (baseline)	**-0.008***	.020	**-0.017*****	<.001	**-0.019*****	<.001
Female			**0.011*****	<.001	**0.016*****	<.001
Age group ** reference group: ≥ 50 years*						
≤ 19 years			-**0.021*****	<.001	**-0.014****	.001
20 – 29 years			-0.006	.167	0.006	.175
30 – 39 years			-0.002	.723	**0.009***	.045
40 – 49 years			-0.008	.054	-0.003	.501
Country ** reference group: Australia*						
United Kingdom			**-0.025*****	<.001	**-0.028*****	<.001
United States			0.006	.060	**0.007***	.024
Canada			0.004	.181	0.001	.726
Education level ** reference group: Higher degree*						
Some secondary			**-0.037*****	<.001	**-0.030*****	<.001
Secondary only			**-0.033*****	<.001	**-0.024*****	<.001
Trade /certificate			**-0.040*****	<.001	**-0.027*****	<.001
Undergraduate			**-0.018*****	<.001	**-0.012****	.001
Diploma			**-0.018*****	<.001	**-0.008***	.029
Bachelor's degree			-0.002	.572	-0.001	.808
Rural/remote location			**0.009****	.003	**0.007***	.011
Depression history			**-0.064*****	<.001	**-0.062*****	<.001
Motivated			**0.060*****	<.001	**0.048*****	<.001
Experience of CBT ** reference group: Not sure what it is*						
Tried, useful			**0.036*****	<.001	**0.026*****	<.001
Tried, not useful			**-0.011****	.001	**-0.013*****	<.001
Not tried, want to			**0.036*****	<.001	**0.026*****	<.001
Not tried, don't want to			-0.005	.077	-0.004	.133
No. treatments			**0.017*****	<.001	0.005	.225
Anti-depressant use			**-0.037*****	<.001	**-0.030*****	<.001

Table B.6. Hierarchical multivariate regression predicting depression change score (n=106,882) (continued).

	Step 1		Step 2		Step 3	
	β	***P***	**β**	***p***	**β**	***p***
No. of modules					**0.134*****	<.001
Time between tests ** reference group: More than one week*						
< 24 hours					**-0.132*****	<.001
1 day to 1 week					**-0.039*****	<.001

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$