CHAPTER EIGHT

General discussion
I

PHOBIAS UNDER CONTROL
INTRODUCTION

The global burden of anxiety disorders is undeniable. Although these disorders – and even more so, phobic disorders – are often underestimated when compared to, e.g., major depression, phobias can seriously affect patients’ lives and can have extensive consequences for a patient’s daily life. What compounds the issue of phobic disorders is a high prevalence, a high comorbidity with other anxiety disorders and low spontaneous remission. Should one have a phobia, one is also probably at risk of developing other mental disorders (Trumpf, Margraf, Vriends, Meyer, & Becker, 2010). Furthermore, phobias rarely come alone and are often comorbid with other disorders such as e.g. major depression (Choy, Fyer, & Goodwin, 2007; Dalrymple & Zimmerman, 2007). Thus, there is a protracted and cumulative effect of phobias on a patient’s well-being and quality of life. Fortunately, anxiety disorders in general benefit greatly from psychological treatment in primary care (Seekles et al., 2013) and in secondary care (Hans & Hiller, 2013). Specifically, psychotherapy for phobias such as social anxiety disorder (Mayo-Wilson et al., 2014), agoraphobia (Sánchez-Meca, Rosa-Alcázar, Marín-Martínez, & Gómez-Conesa, 2010) and specific phobias (chapter 2) all show excellent results.

The advent of Internet-based treatments offered a platform to deliver effective psychotherapies on a much larger scale. An additional benefit was that audiences could be reached that were not previously reached by face-to-face (FtF) psychotherapy, for example adolescents or those in (rural) areas with poor mental health service coverage. When Internet-based interventions were introduced, much was made of the assumed benefits of these interventions in mental healthcare. However, at that time, the soaring expectations for the supposed efficacy and cost-effectiveness for these interventions preceded the available evidence by leaps and bounds; leading to arguably ill-judged implementation practices seemingly at odds with the central tenets of evidence-based practice (as recounted anecdotally in chapter 7 of this dissertation). However, as time progressed, reviews of early trials into the efficacy and effectiveness of Internet-based treatments showed good results ( Andersson & Cuijpers, 2009; Cuijpers et al., 2009), prompting more testing in routine mental healthcare as described in this dissertation. Prior to the
widespread implementation of Internet interventions in routine mental healthcare, and at the onset of this project (2009), an investigation into cost–effectiveness had yet to be undertaken, and the implementation of Internet–based interventions in routine mental healthcare in the Netherlands was limited to a small number of pioneering mental health institutions. The study described in this dissertation attempted to shed light on an attractive benefit of Internet interventions of Internet–based interventions: the possibility of reducing therapist time spent on patients by prefacing FtF psychotherapy with an online intervention. Specifically; the main hypothesis of this dissertation was that an Internet–based intervention with exposure–based homework exercises would be cost–effective in terms of direct and indirect medical expenses, when compared to a waitlist followed by FtF psychotherapy as usual.

SUMMARY OF MAIN FINDINGS

The Internet–based treatment was acceptable and received with enthusiasm by most patients, clinics and clinicians. However, due to methodological limitations described later in this discussion, the main objective of cost–effectiveness was not achieved, although it appears that patients in the intervention group used a statistically non–significant five fewer sessions of FtF psychotherapy (CHAPTER 5) but did not fare substantially worse than the control group participants at 1–year follow–up. Also, a small but statistically significant improvement on the primary outcome in the intervention group was seen at post–test; that is to say at the end of the wait–listed period, after the Internet–based intervention but before scheduled FtF psychotherapy (CHAPTER 4). At 1–year follow–up, this difference between the intervention group and the control group appears to have tapered off, indicating steady improvements for both the intervention and the control groups (CHAPTER 5).
GENERAL DISCUSSION

This dissertation attempted to explore the cost–effectiveness for an Internet–based intervention to reduce therapist time and shorten FtF psychotherapy for phobias in routine mental health care. As detailed in the introduction of this dissertation and in the protocol description of the main trial (CHAPITERS 1 and 3), by making efficient use of the time a patient spends waiting for FtF psychotherapy, we attempted to shorten routine FtF psychotherapy, speed up recovery and reduce both treatment costs and societal costs. Additionally, immediately offering patients some form of treatment could potentially leverage the high motivation for treatment when a patient first presents for treatment. The results of the project will now be discussed.

CLINICAL EFFECTS OF THE INTERVENTION

Some short–term benefit was demonstrated over wait–list in the 5 weeks preceding FtF psychotherapy (CHAPTER 4), where during the relatively short intervention period the participants improved significantly both within group (d=0.42) and when compared to the wait–list control group (d=0.35).

Although it is assumed that spontaneous recovery is relatively rare in phobias (Dingemans, van Vliet, Couvé, & Westenberg, 2001; Green, Hunt, & Stain, 2012), the participants in the control group did improve slightly on the FQ (within–group effect size d=0.13) and modestly on the CES–D (d=0.42); which influences between–group effect sizes at follow–up.

COMPARISON WITH OTHER RESEARCH

Due to the relatively unique setup of this study (an Internet–based intervention during wait–list, followed by FtF psychotherapy), and the relative paucity of data on phobic outpatient samples, there is some difficulty in obtaining relevant comparators for the effect sizes obtained in this study. Most comparable Internet–based programmes in regular clinical settings show higher effect sizes (Andersson & Hedman, 2013), but these are longer in duration as they are not followed by FtF psychotherapy. As such, the most relevant (indirect) comparators are Internet–based and FtF psychotherapy for
anxiety disorders in outpatient settings. The post–test effect sizes (that is to say, those after the Internet–based intervention but before FtF psychotherapy) are lower than in comparable routine mental healthcare samples (Stewart & Chambless, 2009), whose meta–analysis found a mean pre–post effect size of $d=1.04$ (95% CI 0.79—1.29) for 11 face–to–face CBT trials for social anxiety in clinical practice. However, this meta–analysis explicitly excluded studies using fewer than six sessions of psychotherapy, meaning that the results cannot be compared directly to the current study, if some dose–response relationship between treatment and outcome is supposed. A study comparing Internet–based CBT to group CBT for social anxiety disorder found a roughly similar between–groups effect size of $0.41$ (Hedman et al., 2011), however these results are also hard to compare since they include an active comparator rather than a relatively inert wait–list control group. A primary care trial comparing an Internet–based intervention to an “equivalent to care as usual” control group (Nordgren et al., 2014) found an effect size of $d=0.58$, on the BAI, double what our study found ($d=0.28$; chapter 4, table 4.3). However, this intervention also lasted twice as long (10 weeks) as the Phobias under Control intervention. A trial similar in setup and recruitment to the current trial, but focussing on problem–solving therapy for depression (Kenter et al., 2016) found effect sizes even lower than the Phobias under Control trial, with similar dropout figures. For the longer–term outcomes (12–month), we did not show any appreciable difference between the intervention and control groups. As nearly all participants continued to FtF therapy, this attests to the effectiveness of FtF psychotherapy.

The initial differences found at post–test (5 weeks) had, after subsequent FtF psychotherapy equalised (chapter 5, figure 5.2). At 12–month follow–up, the goal was not to demonstrate superiority of the intervention at follow–up, which would be confounded with FtF treatment effects, but to demonstrate that participants were not worse off due to the Internet–based intervention, or a modified FtF intervention due to having participated in the Internet–based intervention. Indeed, following the rationale of the study, one would expect participants to start at approximately similar baseline scores (due to randomisation), diverge in clinical scores immediately after five
weeks (due to a treatment effect), and end at approximately similar follow-up scores (due finishing a FtF treatment trajectory); but with intervention participants needing fewer sessions of FtF psychotherapy. Indeed, patients in the intervention group improved quicker in the five-week period than patients on the wait-list and could start working on their issues immediately, possibly boosting motivation and empowerment (CHAPTER 4). However, it is possible that treatment was not ‘finished’ for some participants at the 12-month mark.

The clinical effects were analysed on a conservative intention-to-treat basis (Hollis & Campbell, 1999), which is considered to be the conservative approach. By analysing all those allocated to an intervention (even if they did not receive this intervention), intention-to-treat analyses are justifiably seen as the hallmark of an adequately analysed RCT, but clinical practice is not bound by these restrictions. The intention-to-treat analysis paradigm has also been criticised as it rests on assumptions that may implausible given specific trial circumstances (White, Horton, Carpenter, Statistics, & Pocock, 2011). In regular psychotherapeutic practice, skilled therapists constantly adapt and tailor therapy to individual patients based on their skills and preferences. In clinical practice the therapist will attempt to prevent treatment discontinuation, or try to get the patient back into treatment – in short, therapists and patients are interested in, and prefer treatment courses that work for individuals (Batterham & Calear, 2017), not for a hypothetical average patient in a mixed set of patients included in a RCT.

Thus, given the low adherence to the investigated intervention, an intention-to-treat analysis may underestimate the true treatment effect in clinical practice, since patients that actively choose an Internet-based intervention are probably more likely to complete it and to derive clinical benefit. Moreover, there is some evidence that it is beneficial if patients are allowed to choose which modules of an Internet-based intervention seem most appropriate (Andersson, Estling, Jakobsson, Cuijpers, & Carlbring, 2011). As such, it could be argued that the clinical effects of the intervention were tested in what could be described as ‘worst case’ circumstances if the intervention is indeed effective.
There was a very high occurrence of comorbidity in the sample of this project—although we administered just the anxiety and depression subdomains of the CIDI diagnostic interview (World Health Organization, 1990), it would be more accurate to describe the sample in terms of multi-morbidity. The mean number of phobias in the sample was around 2, with 30% of participants having three or more phobias (Chapter 4, Table 4.1). Comorbid GAD and MDD were also common (Chapter 6, Table 6.2), with the mean CES-D of the entire sample at baseline being 24.8 (Chapter 4, Table 4.1), well over the commonly accepted cut-off score of 16 for elevated risk of depression (Lewinsohn, Seeley, Roberts, & Allen, 1997). To all intents and purposes, the current sample might as well have been included in a depression trial.

A register-based survey of Dutch patients seeking outpatient help for social anxiety (Dingemans et al., 2001) showed that 50% of patients suffered from one comorbid axis-I disorder, and 35.9% of patients suffered from two comorbid axis-I disorders. In the current sample, however, these comorbidity figures were much higher: 78.1% of patients had two or more axis-I anxiety or depression diagnoses, and 48.6% had three or more. Even when limited to anxiety and depression, the sample in the Phobias under Control project, in short, was quite unwell.

The question is, then, whether a short, Internet-based intervention is something that would be useful for a patient sample with severe comorbidity, even if it is followed up by FtF psychotherapy. If patient characteristics are such that it is likely that patients will discontinue the intervention, or derive very little benefit from it, then it is rather useless to offer the intervention. In fact, a survey among clinicians suggests that a diagnosis of anxiety makes it less likely that clinicians will recommend a brief (six sessions or fewer) treatment (Schaefer, Koeter, Wouters, Emmelkamp, & Schene, 2003), and in the current sample, baseline anxiety was a predictor of nonadherence to the intervention (Chapter 6, Table 6.5). However, seeing clinical levels of symptom severity in Internet-based treatments is not rare (Titov, Andrews, Kemp, & Robinson, 2010), which indicates that despite symptom severity,
patients see Internet–based treatments as a viable alternative (or adjunct) to FtF psychotherapy. However, many of these interventions originate from low–intensity self–help interventions, intended for subclinical or moderate symptoms, and one cannot help but think that these converting those interventions to the Internet and offering them to clinical patients is simply an example of ‘the wrong tool for the job’. Despite the motivation of these patients, adherence is lacking, and the challenge is to make the interventions relevant, engaging and useful for patients with severe symptoms.

**COST–EFFECTIVENESS OF THE INTERVENTION**

Since the start of the trial described in this dissertation, evidence for the effectiveness of Internet–based treatments in both the general population and general healthcare began to emerge (Andrews, Cuijpers, Craske, McEvoy, & Titov, 2010; Richards & Richardson, 2012), and gradually during the course of our evaluation, tentative indications for cost–effectiveness outside clinical settings were underway (Donker et al., 2015). Indeed, the presupposed cost–effectiveness is often brandished as a reason for implementing Internet–based interventions in routine mental healthcare (Chapter 7). However, in the early 2010–s, a full economic evaluation into Internet–based treatments in routine mental healthcare had not yet been undertaken to the best of our knowledge. The intervention was relatively inexpensive, at a cost of €252 per participant. Since FtF sessions were valued at €180.50† per session (Chapter 5); in clinically interpretable terms this means that the five–week intervention cost roughly 1.5 sessions’ worth of FtF psychotherapy. At 12–month follow–up, the intervention group used 4.8 fewer sessions of FtF psychotherapy than the control group (23.4 sessions in the 12–month period in the intervention group and 28.2 in the control group during the same period). This translates into a total cost of €866.40, if clinic costing is taken into account. The intervention group also incurred slightly lower costs on a societal level, but this difference was not significant. For the greatest part, the reduction in costs was due to the lower number of FtF sessions used by the intervention group. The total costs on a societal level for the intervention group were €481 less

† This costing method includes all operational costs of one hour of therapy.
than for the control group, but this difference was not statistically significant, nor was this significant for the per-protocol analyses, where a societal cost reduction of €1525 appeared (CHAPTER 5, TABLE 5.4).

It should however be said that the analyses were not powered to find a difference in costs, and given the small effects of the intervention, it would be nearly impossible to find sufficient participants for a well-powered economic evaluation (Al, van Hout, Michel, & Rutten, 1998; Briggs, 2000). Notwithstanding this, there are encouraging indications for reduced FtF session uptake. As cost-effectiveness is – not unsurprisingly – calculated using both costs and effectiveness of an intervention, the standard interpretation of cost-effectiveness research is summed up by the incremental cost-effectiveness ratio (ICER).

For the primary outcome measure (FQ) in the intention-to-treat analysis, the ICER was 166 (491 for the per-protocol sample), meaning that a 1 lower point on the FQ was associated with €166 (or €491) lower costs. The probability that the intervention would be cost-effective compared to the control group was 0.65 — but only at the willingness-to-pay point of €0. In essence, there would be a 65% possibility that the intervention would be cost-effective if nobody is prepared to pay for it (CHAPTER 5, FIGURE 5.3).

Comparing these results to those synthesised by a systematic review of economic evaluations of Internet-based interventions (Donker et al., 2015), the Phobias under Control intervention performs worse than the anxiety interventions reviewed in that review. Compared to the ‘sister’ trial for depression mentioned earlier (Kenter et al., 2016), the economic results were similar, but with higher societal costs for the intervention group (Kolovos et al., 2016).

Ultimately, the consideration of whether the smaller clinical effect is offset by the lower costs is a decision that cannot be made objectively. It depends on what clinics are willing to sacrifice in terms of effectiveness in return for slightly reduced costs. Moreover, this decision has to take place against a backdrop of secondary and unquantifiable benefits and drawbacks, such as the ability to keep patients engaged prior to FtF treatment (benefit), or having to change the workflow of the clinic (drawback).
ACCEPTABILITY OF THE INTERVENTION

Adherence

Apart from the recruitment of clinics and patients, a second issue is the lack-lustre adherence to the intervention, which is all too common in the field of psychotherapy (Santana & Fontenelle, 2011) and Internet–based interventions (Christensen, Griffiths, & Farrer, 2009; Neve, Collins, & Morgan, 2010; van Ballegooijen et al., 2014). As also demonstrated in a randomised controlled trial similar to the one described in this dissertation (Kenter et al., 2016), it was extremely difficult to encourage participants to continually engage with the intervention despite weekly messages from the coach. Moreover, there appeared to be a connection between not completing the Internet intervention, and not completing follow–up questionnaires (CHAPTER 6), which will be explored in the following section. At post–test, loss–to–follow–up was 46.7% for the intervention group and 32.7% for the control group. At 1–year follow–up, these figures had increased to 59.0% and 54.2%, respectively (CHAPTER 5). Differential loss–to–follow–up may be an issue if there is an imbalance in dropout between the intervention and control group (Graham & Donaldson, 1993).

It has been speculated that participants in the intervention group may have higher expectations of the intervention, leading to lower questionnaire compliance (Crutzen, Viechtbauer, Spigt, & Kotz, 2014); but differential attrition may occur because of myriad reasons which are not necessarily connected to the trial itself; even trivial reasons such as not being able to access an e-mail address or simply not having the time or the motivation to complete the assessment. Additionally – and anecdotally – it had come to our attention that some patients had confused clinic–provided, treatment–related assessments from routine outcome measurements with our university–provided, research–related assessments; leading patients to think that they had already completed the assessment for which reminders had been sent.

Although the loss of participants to follow–up is often accounted for in sample size determinations, the estimate of 30% dropout from follow–up measurements in the RCT (CHAPTER 3) to hedge against loss–to–follow–up
was, in hindsight, too low. However, recruiting more patients would be both impractical and wasteful – with 50—60% measurement dropout, every second recruited patient would be lost to follow-up. Even though some attempts have been made to reduce loss to follow-up from Internet intervention trials, the impact of variables such as reducing questionnaire length or participant remuneration have not led to any conclusive strategies for reducing questionnaire dropout (Khadjesari et al., 2011; McCambridge et al., 2011). However, there are opportunities in the clinics’ routine outcome measurement system (ROM; de Beurs et al., 2011; Chapter 7) which will be discussed later on. It is difficult to pinpoint the reason for the low adherence figures, and previous research (Karyotaki et al., 2015; van Ballegooijen et al., 2014) has mostly focussed on clinical predictors or baseline statistics and demographic information to predict adherence to Internet-based interventions (such as Chapter 6).

However, one pragmatic reason to be considered specifically for the Phobias under Control trial is that the FtF psychotherapy directly following the Internet-based intervention could have lowered the threshold for participants to discontinue the intervention. All patients were ensured of FtF treatment and were explicitly informed that not participating in, or discontinuing, the Internet-based treatment would have no consequences for their FtF treatment. Randomised controlled trials into Internet-based therapies have often employed a ‘one choice only’ design, where the Internet-based therapy was the only treatment offered to patients; other, less common designs have used non-inferiority or equivalence trials, where participants are randomised to only one of two or more options. However, the trial design that we have researched offered two concurrent treatment options to patients, which put patients at liberty to discontinue the one treatment in lieu of the other treatment without being left without any treatment at all.

PARTICIPANT MOTIVATION

It should be noted that patient motivation to enrol in our trial was relatively high; and 68.2% of patients we initially contacted agreed to participate. Ultimately, our trial randomised 212 out of 481 referred patients (44.1%), showing enthusiasm for the treatment format. Of these 481 participants, 153
declined to participate, which includes those patients averse to the idea of possibly being randomised into a wait–list condition (CHAPTER 4, FIGURE 4.1). This is relevant, and an important detail is that all contacted patients were unaware of both the trial taking place and the Internet–based intervention until they were informed of this by outpatient staff. It is not unthinkable that even more patients would have been interested in the Internet–based intervention if this had not been combined with a randomised controlled trial and its associated burden and inconveniences for the patients. Finally, 111 patients (23.1%) did not meet inclusion criteria (CHAPTER 4, FIGURE 4.1). It should be noted that all patients possibly meeting inclusion criteria were referred to the research team to reduce the risk of missing patients (CHAPTER 2); therefore, it is not surprising that 23.1% of patients did not meet formal inclusion criteria.

The initial patient interest in the project is encouraging, and serves perhaps to underline the high motivation for treatment at that point in the patient’s treatment path. This is a marked difference with trials that recruit patients using self–referral, or recruit from the general population by way of newspaper advertisements; where self–selection by patients highly motivated for, and interested in, specifically Internet–based interventions can be an issue. It is encouraging that fewer than 1/3rd of patients expecting FtF psychotherapy declined to participate in a research project which not only introduced a new form of therapy, but also involved being randomised and filling in additional questionnaires. This positive result was found previously in an outpatient clinic, where 52.8% of patients preferred starting with an Internet–based intervention rather than wait for FtF psychotherapy (Kenter, Warmerdam, Brouwer–Dudokdewit, Cuijpers, & van Straten, 2013). An additional potential effect that our Internet–based intervention attempted to leverage was the motivation for treatment of patients. It is not inconceivable that motivation drops during the wait–list period, or that patients drop out before even starting FtF treatment — one study estimated that 30.5% of outpatients awaiting FtF psychotherapy never started treatment after the wait–list (Issakidis & Andrews, 2004). It has even been suggested that patients are ‘motivated’ to stay ill during a wait–list condition (Furukawa
et al., 2014), although other research has seen improvement during wait-list for the treatment of anxiety disorders (Young, 2006).

Moreover, it should be noted that, in the Dutch system, a person presenting at an outpatient clinic for treatment has already been seen by a general practitioner; yet it is generally unknown which proportion of patients referred to outpatient clinics actually present themselves for treatment. That is to say, patients presenting themselves for FtF treatment at an outpatient clinic have already passed several hurdles in obtaining treatment. Passing such hurdles must require considerable motivation for treatment, therefore it is of importance that patients that have proceeded to FtF psychotherapy are not discouraged or demotivated by the presence of a wait-list.

PATIENT PREFERENCES AND EXPECTATIONS

It stands to reason that patients presenting at an outpatient clinic for FtF psychotherapy expect FtF treatment format rather than an Internet–based therapy; and all but four patients proceeded to FtF psychotherapy after the Internet–based intervention or the control group (Chapter 5). The ideal situation would be one where all patients follow the entire Internet intervention. However, in a situation where patients will receive FtF psychotherapy anyway, one could argue whether it is inherently undesirable for patients to prematurely discontinue an intervention. In our trial, no patient went without treatment, and ultimately all but four patients continued to FtF psychotherapy. Most patients had improved at 1–year follow-up (Chapter 5, Figure 5.2); and regardless of which treatment path patients took, this can be taken as a success for the patients and for the clinics.

Although it has been argued that Internet–based interventions have been oversold (Chapter 7), it could equally be argued that the intervention in this trial was undersold from the perspective of routine healthcare. At face value, the outpatients in this trial were being offered, free of cost, a voluntary enrolment into a programme explicitly described as experimental, and being offered by a university rather than an outpatient clinic. One could easily imagine such a proposition not exuding a great deal of confidence in the offered intervention, and in hindsight one must wonder what sort of
impression the intervention (and associated trial) must make on interested patients. Previous research has suggested that treatment credibility can be a factor in psychological treatments (Mooney, Gibbons, Gallop, Mack, & Crits-Christoph, 2014) as well as in Internet-based interventions (Alfonsson, Olsson, & Hursti, 2016); and patient expectations appeared to predict better adherence and better outcomes in a trial of Internet-based therapy for social phobia (Boettcher, Renneberg, & Berger, 2013). Unfortunately, treatment credibility was not measured in the context of the current trial. Had the intervention been confidently offered as a regular part of routine healthcare, then perhaps patients would have accepted and adhered to the Internet-based treatment as a regular, inseparable part of routine treatment.

In conclusion, it could be said that the patients enrolled in this trial were willing and motivated to follow the Internet-based intervention, despite the fact that they had no knowledge of the intervention until after they were placed on a wait-list for FtF psychotherapy. However, the low adherence to the intervention also shows that having started the intervention, many patients did not seem to be overly enthusiastic about the format, delivery or content of the intervention, and a more detailed analysis of patients’ appraisal of interventions prior to testing it in a RCT can be worthwhile. This may be done with e.g., focus groups, or with pilot testing, the latter of which will be briefly discussed later.

EMBEDDING INTO ROUTINE MENTAL HEALTHCARE

The popularity of Internet-based treatments had, at the start of this trial, not gone unnoticed to most outpatient clinics, and fortunately many of the clinics were keen to cooperate. However, despite best efforts from all parties involved, many administrative and logistic hurdles delayed the effective and timely recruitment from the clinics. Internal approval procedures for ethical and legal requirements proved to take longer than anticipated, and a number of interested clinics indicated that they simply did not have the resources to support running a randomised controlled trial. This demonstrates that embedding research in routine mental healthcare is not always feasible. Seeing these impediments, it is a considerable credit to outpatient clinics involved
in this trial that they were willing to expend time, staff and resources to a complex undertaking that has no immediate, tangible benefit for the clinics.

A recent survey among professionals found that issues raised in the implementation of Internet–based interventions were not unique to the Internet format, but they were broader, more general concerns such as access to care and stigma (Nicholas et al., 2017). Thus, it is questionable whether such arguments would be valid to conclude that Internet–based interventions specifically are poorly suited for implementation in routine mental healthcare. Nonetheless, a number of pragmatic issues encountered while embedding the Phobias Under Control trial in outpatient clinics will be briefly discussed.

Embedding a randomised controlled trial in routine mental healthcare

The preparation and running of a large–scale randomised controlled trial in outpatient settings requires clinics to make considerable changes to broad range of routines, such as enrolling patients and diagnosing them. Furthermore, there is a considerable increase in administrative duties on top of existing administrative duties (legally) required for the daily management of a clinic. In fact, it could be said that the clinic is asked to keep an entire parallel administration to be able to enrol and keep track of patients enrolled in a randomised controlled trial; which is especially taxing since the administrative workload of most clinics is very high as it is.

The issue of administration is often compounded by IT–systems (such as the electronic health record) which are, understandably, primarily tailored to meet the administrative demands of providing FtF psychotherapy to patients; and not to perform research on Internet–based therapies. As an example, time spent by outpatient staff to facilitate the trial (e.g. communicating with researchers) could not be entered as reimbursable time in some systems; amounting to a net loss in revenue to the clinic. This financial issue can sometimes be mitigated by external funds dedicated to research, but not all clinic expenditures can be offset by money alone. Business management in outpatient clinics is increasingly confronted the expectation of treating more patients with fewer funding. It is therefore essential that clinics optimise and streamline procedures and workflow. However, conducting a randomised
controlled trial can be a disruptive experience for clinics, since it requires changing these existing procedures and workflow in the clinic to facilitate the extra trial–related actions described earlier. Again, it is to the great credit of outpatient clinics and staff who were motivated enough to go through considerable effort to enable this trial; especially given that the disruptive logistical and methodological demands of the trial necessitated creative and original solutions to circumvent the inflexibilities of some of the outpatient clinics’ systems and operational procedures.

Recruitment of outpatients

The first issue encountered in the RCT described in this dissertation is that the accrual of patients was an arduous process, and that the number of recruited patients was ultimately insufficient. Although the trial originally sought to enrol 320 patients in one year, we managed to include only 212 patients in three years of recruitment — effectively reducing the number of recruited patients by 1/3rd while overrunning the planned time span needed by a factor of three. Although those of a more sunny disposition would point to similar trials that fared even worse (Woodford, Farrand, Bessant, & Williams, 2011); there is no question that the enrolment of patients with complex mental health issues from specialised mental health clinics has considerable organisational implications, such as described in the previous section.

The complexity of enrolment is exacerbated when trying to not only enrol patients into an Internet–based treatment (while patients were expecting FtF treatment) but simultaneously trying to enrol patients into a randomised controlled trial. Low recruitment rates to randomised controlled trials are common, with a reported less than one third of trials recruiting the target number of participants in time (Campbell et al., 2007). For some patients, the combination of a novel therapy coupled to the perceived ‘oddity’ of being enrolled into a RCT may be too much hassle to participate.
CONCLUDING REMARKS

Two issues are relevant as a conclusion to the main effects of the intervention. Firstly, as found previously (Kenter et al., 2013), the Internet–based intervention is acceptable and appealing to outpatients in routine mental healthcare who are waiting for ‘traditional’ FtF psychotherapy.

Secondly, some clinical effect is found. Although cost–effectiveness was not demonstrated, the intervention was relatively cheap (per patient, the costs were approximately those of 1.5 hours of psychotherapy); moreover, patients were introduced to the principles of exposure treatment and encouraged to build self–efficacy. With regards to clinical practice, when research–related restrictions such as randomisation and exclusions are removed, these interventions can be opened up to a larger group of patients, and can be a good way to activate and empower patients, and to prepare them for upcoming FtF treatment.
LIMITATIONS

STUDY DESIGN

Although key limitations to this main study have been discussed previously in respective chapters of this dissertation, some issues and considerations related to the main RCT of this dissertation are explored in more detail in the following paragraphs. Broadly, the two most salient methodological issues of this dissertation are related, yet distinct in possible causes and plausible effects, being the RCT as a study design, and study dropout. Firstly, the RCT is, and will probably for a long time, remain the gold-standard in effectiveness research, prompting to use this design in the current outpatient clinic setting. Secondly, the issue of dropout has implications for statistical power.

Although a formal sample size calculation was performed prior to this trial (CHAPTER 3), a priori sample size estimations of RCTs are essentially synthetic, statistically driven estimates of the number of participants needed to achieve sufficient statistical power to reject a null hypothesis. This requires a good deal of informed guessing and assuming on the part of the researchers†, including (but not limited to) the expected effect size. However, in relatively new settings, this effect is generally not known, and based on relatively scarce previous research. This led to a situation where for years, effect sizes for sample size calculations for Internet-based interventions usually came from a single meta-analytic source (Spek et al., 2007). However, this source included mostly non-clinical populations and might be quite inappropriate in clinical settings. In hindsight, the estimated effect size of \( d=0.7 \) (CHAPTER 3) for the intervention group for our RCT was twice that of the actual short-term effect size of \( d=0.35 \) (CHAPTER 4, TABLE 4.3). The large loss-to-follow-up (33–47% at post-test; 50–60% at 1-year follow-up) necessitated the statistical imputation of data to mitigate the impact of missing data; however, there is only so much a researcher can do when faced with the issue of missing data. Therefore, we opted for a pragmatic approach of multiple imputation with predictive mean match-

† Voodoo is also rumoured, but this is usually dismissed as malicious truth.
ing at item–level rather than the total score level (Chapter 5), as per recent recommendations (Eekhout et al., 2014). However, the large percentage of missing data is a caveat for the interpretation of the results. It should be noted that care must be taken in terms of the generalisability of the results in the light of this limitation.

LIMITATIONS — CONCLUDING REMARKS

When summarised, these limitations may paint a frankly less than encouraging picture for the intervention and its execution — taking too long to recruit too few participants who largely discontinued an intervention that was less successful than anticipated. Anyhow, one bleak conclusion could be that the trial was underpowered and inconclusive, despite statistically significant short–term outcomes. Regarding recruitment, ultimately, it is hard to predict potential incident cases at outpatient clinics; which necessitated making informed guesses about the potential number of patients based on population prevalence estimates and estimates from clinics themselves. Given that phobias are often reported as secondary to, e.g., major depression, the potential number of new patients was surrounded by a great deal of uncertainty.

Although a long recruitment period is not in itself an issue, a prolonged recruitment period slows the generation of evidence for a fast–moving field; and it places an undue strain on outpatient resources. However, when viewed in the historical context – one must allow that five years is a substantial amount of time in anything Internet–based† — important and valuable information was retrieved about research in routine mental healthcare. Much of this information, though not formally documented in published articles, was subsequently re–used in other projects recruiting patients from similar settings (Kooistra et al., 2014; Romijn et al., 2015).

Possible solutions for a number of the limitations observed in this section will be discussed in the sections dealing with implications for research and implications for clinical practice.

† For example, the rise and fall of the immensely popular, yet ill–fated MySpace platform took place in just over five years.
IMPLICATIONS FOR CLINICAL PRACTICE

Following the previous sections, there may be good reasons for outpatient clinics to implement Internet–based interventions, even those that are not evidence–based†.

These treatment options should be carefully discussed with the individual patient. Internet–based interventions could be a valuable tool in the therapist’s toolkit, yet the decision to use these interventions should be guided by individual patient preferences, therapist recommendations and solid guidelines; the latter of which are currently lacking. These guidelines would ideally be a collaborative effort between clinicians, patients and researchers to determine the most productive scenarios for the use of Internet–based interventions. It is unlikely that the number of patients will decrease in the future — perhaps prevalence figures will not change, but a growing population and increased acceptance of psychological treatment will lead to an increased demand for psychological treatments. Sadly, it is unlikely that the increasing burden on mental health services will also lead to a commensurate increase in mental health budgets.

As indicated earlier, the use of Internet–based interventions can be a valuable tool for clinics, alongside services such as FtF psychotherapy. In view of the results from the Phobias under Control project, while keeping in mind the limitations described in this dissertation, a number of arguments can be made to encourage the use of an Internet–based pre–treatment in routine mental healthcare.

Firstly, the motivation and subsequent acceptance of the intervention was high in a representative sample of non–self–selected outpatients.

Secondly, the intervention brings about positive clinical effects in the relatively short time–span of five weeks. Thirdly, due to the low amount of guidance needed for the intervention, the intervention could be offered at a modest cost per patient.

Fourthly, there are indications that the intervention group, as hypoth-

† There are also bad reasons to implement non–evidence–based interventions, as described in CHAPTER 7.
esised, used fewer sessions of FtF psychotherapy after the Internet–based intervention, although this difference was not statistically significant.

Fifthly, with some tenacity and inventiveness (and a lot of cake–based motivational interventions for healthcare staff), it was possible to embed the intervention – and accompanying RCT – in routine mental healthcare, demonstrating the feasibility of using such an intervention in real life.

Sixthly, all but one patient reported no adverse events during the Internet–based intervention, and both the intervention group and control group had comparable scores at 12–month follow–up, demonstrating that when the entire treatment trajectory is taken into account, patients who participate in the Internet–based intervention are not worse off than patients in the control group.

Finally, according to a recent survey, the average time spent waiting for FtF psychotherapy for adults in the Netherlands is still five weeks (MediQuest, 2017); identical to the wait–list period when this project started (2009). Given that the wait–list times are unlikely to go down, outpatient clinics could use an Internet–based intervention to give immediate attention to patients; and educate, activate and empower patients as a preparation for FtF psychotherapy.
II

REFLECTIONS &
FUTURE DIRECTIONS
A relatively recent development attempts to combine the best characteristics of both FtF and Internet–based interventions: blended treatments. Although some research efforts are underway (Kleiboer et al., 2016; Kooistra et al., 2014; Mathiasen, Andersen, Riper, Kleiboer, & Roessler, 2016; Romijn et al., 2015), the current paucity of evidence on blended treatments means that there is no way of knowing what the optimal blend of FtF and Internet–based therapy is.

On top of this, preliminary evidence from interviews suggest that therapists and patients have differing opinions on the matter, with therapists wishing a predominantly FtF intervention, but patients wishing a predominantly Internet–based intervention (van der Vaart et al., 2014). However, this was a very small sample of patients, and the flexibility of Internet–based interventions means that this mix can be tailored to individual patients as needed.

It is interesting and relevant to speculate on whether the intervention described in this dissertation would have been more successful in a blended setting. Although the issues pertaining to the RCT would remain the same, perhaps the issues of low adherence to the intervention could have been mitigated by involving the therapist in the Internet–based treatment. Our Internet–based treatment was offered as explicitly distinct from the FtF psychotherapy. By design, therapists were informed of inclusion status into the trial, informed on whether the patient had actually completed the Internet–based intervention, but not informed on specific content in the Internet–based intervention.

This procedure was designed so as to allow the therapist to tailor the FtF intervention to allow for treatment ‘as usual’, that is to say, to allow the therapist to assess the patient’s existing knowledge on, e.g., the rationale of exposure therapy. This way, the therapist could change the content, form and length of the intervention to make use of what the patient had learnt and done during the Internet–based intervention. In blended therapy, such a procedure would take place every FtF session, possibly helping to consolidate the treatment content and helping the patient to adhere to the online modules.
As it stands, the current evidence on blended Internet–based interventions is too scarce to warrant any conclusions either way. However, if one is to assume that, evidence pending, blended will combine the best of both worlds, one must equally allow that – mutatis mutandis – it can combine the worst of both worlds. For example, Internet–based therapies allow rural populations to receive help in remote locations – but this advantage is negated when in blended treatments the patient again needs to see a therapist FtF. As research results on blended therapies slowly accumulate, the place of blended therapies in the mix of treatment options will hopefully soon be established.

IMPLEMENTATION OF INTERNET–BASED INTERVENTIONS

Since the start of this project, many outpatient clinics have enthusiastically set out on a process of implementing Internet–based interventions. The number of Internet–based services on offer in the Netherlands has positively mushroomed since the inception of the Phobias under Control project. Although such enthusiasm is commendable and encouraging, it seems that many implementation initiatives have been haphazard, and driven by assumed benefits and evidence that have yet to be confirmed by rigorous scientific evidence (Chapter 7). Many Internet–based interventions have been adapted from earlier, existing self–help books which are based on cognitive behavioural therapy (CBT; Chapter 7). Although self–help can be effective (Haug, Nordgreen, Öst, & Havik, 2012) and self–help with guidance from a professional can be as effective as face–to–face psychotherapy (Cuijpers, Donker, van Straten, Li, & Andersson, 2010), many Internet interventions currently in circulation have been derived from self–help interventions, but with considerable changes from the original book. Careful observation will reveal that books, on the whole, are overwhelmingly text–based†, but for the Internet these self–help texts are often shortened, rewritten or paraphrased. However, there are no set guidelines or rules for what to therapeutic content retain and what to discard, with those doing the editing left to choose which elements of the therapy they

† Excluding those targeted at the under–threes.
deem to be effective. Thus, not only the content but also the presentation of these interventions is altered substantially. Arguably, one could say these are new interventions altogether, and need research into their efficacy or effectiveness (CHAPTER 7).

Since originally these interventions were based on the idea of, e.g., CBT, an ‘evidence-based’ status is appropriated to these interventions where perhaps ‘evidence-assumed’ is more appropriate. A point is to be made that these derivations of interventions have neither been assessed in their current form and current presentation, nor in their current populations or clinical settings. Conversely, interventions that have taken years to be rigorously assessed remain as yet unimplemented, perhaps partly because they have been displaced by the readily available (yet unassessed) interventions available from commercial parties offering Internet-based treatment solutions (CHAPTER 7).

Some initiatives have started to streamline the implementation of Internet-based interventions, e.g. the ImpleMentAll initiative as recently funded by the European Commission (grant number 733025). This project uses the MAST model (Kidholm et al., 2012), which is an assessment framework developed specifically to assess the prerequisites for implementing technology-based interventions in healthcare. It assesses domains such as legislation, patient perspectives, safety, organisational aspects and other important factors that are deemed to be determinants of implementation success. Analogous to this MAST-framework, which is used before a technology is implemented, the RE-AIM (Glasgow, Vogt, & Boles, 1999) framework is currently being used to assess factors that are associated with a successful uptake of new technologies, such as reach, adoption, efficacy and maintenance.

For example, one project is currently examining therapist uptake of blended Internet-based interventions in routine mental healthcare (Mol et al., 2016). Other, new toolkits for determining what kind of blended treatment to implement have also been developed (Wentzel, van der Vaart, Bohlmeijer, & van Gemert-Pijnen, 2016), paving the way for clinics to base their implementation strategies to best practices developed by others.

At any rate, the support of therapists is very important for successful implementation of Internet-based interventions in routine mental healthcare.
Since resistance to change in organisations can cause implementation projects to fail completely and cause considerable stress among employees (Prochaska, Prochaska, & Levesque, 2001), the support of all organisational layers is needed for such projects to succeed. Next to organisational support, adequate business modelling (van Limburg et al., 2011) could help to lay a robust foundation for sustainable and cost-effective Internet-based interventions, even before the question of cost-effectiveness is addressed. The implementation of an Internet-based intervention can mean a considerable investment for clinics, both up-front in terms of the purchase of software, and recurrent in terms of subscriptions and/or maintenance. Since cost-effectiveness data from outpatient clinics is currently scarce and not unequivocally promising (Donker et al., 2015; Kolovos et al., 2016; Chapter 5), clinics would do well to research whether this is an investment that will actually pay off when implemented in routine mental healthcare, even with financial incentives in place (Chapter 7). At the end of the implementation, it can be extremely difficult to un-implement an intervention should it prove to be ineffective or not cost-effective (Prasad, Cifu, & Ioannidis, 2012). Though arguably less methodologically robust, there is a wealth of knowledge within clinics that can be tapped into: routine outcome measurement (ROM) data (Chapter 7).

At its core, a clinic’s primary interest is to ensure adequate and effective treatment for all patients, and though lacking a control group, something is to be said for a relatively simple monitoring of patients, where routinely collected clinical data is used to assess at least a patient’s trajectory throughout a treatment, as is already partially in place in at least the Netherlands (de Beurs et al., 2011). Valuable comparisons can at least be made when the clinic has wait-listed patients. Although it has been suggested that Internet-based interventions can actually cause harm (Rozental et al., 2014), Internet-based interventions could be used by clinics as a low-risk stop-gap before FtF treatment. This is especially salient when offered to motivated patients by enthusiastic and supportive staff; and possible negative effects could be monitored by way of short, routine check-ups with the patient. Thus unencumbered by all the logistical and methodological restrictions of a RCT, interesting data could be retrieved from a realistic outpatient clinic setting.
There remains the ethical issue of offering an intervention which is, at least to some standards, probably not evidence-based; but as patients are to be offered FtF psychotherapy anyway, this decision is at the outpatient clinic’s discretion. Crucial, then, as in stepped-care treatments (Bower & Gilbody, 2005), is the careful monitoring of patient clinical status, wishes, needs and preferences. Ultimately, the clinics are at liberty to offer Internet-based interventions without the randomisation procedure required in RCTs. In this way, these interventions can be offered solely to motivated patients who have expressed a keen interest and are willing and able to do the intervention, selecting those patients where a good effect is to be expected. This can prove to be very valuable in routine practice; and research into what works for whom would help to indicate which patients especially benefit from these interventions (Ebert et al., 2013). With more insight for whom these interventions are particularly helpful, Internet-based interventions can become an indicated tool in a clinic’s toolbox of therapies.

Implementation — concluding remarks
The implementation of Internet-based interventions in routine mental health-care will, for better or worse, continue unabated in the future, regardless of academic opinion. As noted previously, de-implementation of ineffective interventions can be difficult. Therefore, sufficient, clear and honest dissemination of previous (negative) experiences is of vital importance to help clinics avoid previously encountered difficulties. Internally, practitioner-scientists (CHAPTER 7) can help streamline the implementation and attempt to make the most out of the available data.

There is no doubt that routine mental healthcare clinics tirelessly work to improve and expand patient-centred care using the best and most desirable tools available; and researchers should strive to help clinics wherever they can with guidance, advice, practical support, and accessible scientific publications with clear implications for clinical practice. A fruitful collaboration with the joint ultimate aim of providing evidence-based, patient-centred Internet-based interventions is certainly possible; and some implications for clinical practice are described earlier in this discussion.
Trial design

As noted in previously, clinical and technological reality seem to outpace research into Internet–based interventions. The quest is then to find alternatives to traditional RCTs that are better suited to Internet–based interventions, while retaining methodological rigour. Currently, it can be foreseen that future RCTs into Internet–based interventions will increasingly have to test effectiveness against treatment–as–usual or other therapeutic comparators. This would mean an increased occurrence of non–inferiority and equivalence trials (Greene, Morland, Durkalski, & Frueh, 2008); but these require a larger sample size than traditional superiority trials (Lenth, 2001). A number of alternatives to traditional RCTs have been proposed, which will briefly be examined with an eye to suitability for Internet–based interventions. The first alternative to the ‘classical’ RCT for Internet–based interventions could be a stepped wedge cluster RCT design (Hemming, Haines, Chilton, Girling, & Lilford, 2015), which shows promise in complex health systems (Keriel–Gascou, Buchet–Poyau, Rabilloud, Duclos, & Colín, 2014) since it randomises at the cluster (viz., clinic) level rather than the individual level. The design of such a trial would fit the gradual, ‘rolling’ implementation of an Internet intervention in a healthcare system.

However, it is not completely sure whether this design is actually preferable to a ‘normal’ cluster RCT (Kotz, Spigt, Arts, Crutzen, & Viechtbauer, 2013). Furthermore, since cluster RCTs and stepped wedge trials tend to require larger sample sizes to obtain similar statistical power to regular RCTs (Campbell, Elbourne, Altman, & CONSORT group, 2004; Hemming et al., 2015), it raises the question whether the additional complexities of these designs outweigh the benefits over non–inferiority trials. However, one very compelling advantage, especially in outpatient psychotherapy, is that the stepped wedge design can make effective use of ROM data (Hemming et al., 2015), which is increasingly gaining traction even in less research–minded outpatient clinics (de Beurs et al., 2011). Another candidate would be a regression discontinuity (RD) design
(Bor, Moscoe, Mutevedzi, Newell, & Bärnighausen, 2014), a quasi-experimental, nonrandomised approach to interventions administered on a large scale. In short, the RD approach assigns patients to conditions based on cut-off scores, as often the case in the field of Internet-based interventions. As the RD design presupposes less strictly controlled environments than RCTs, there may be a trade-off where some internal validity (high in RCTs) is traded for higher external validity, as participants in a RD study are not randomised and excluded based on certain exclusion criteria. The RD analysis was recently successfully applied — albeit, post-hoc) to data from a RCT on an Internet-based intervention (Maas et al., 2017), showing some promise as an alternative approach to RCTs on Internet-based interventions. However, it may take a considerable amount of time before these alternative designs are validated and accepted as an alternative to the ‘gold-standard’ RCT, and some methodological purists might even insist that this is impossible†.

Whatever newer, more pragmatic research design – if any – finds acceptance is hard to predict, and surely it is conceivable that like psychology, the fields of statistics and epidemiology are also susceptible to fashions, hypes and trends. At any rate, a preoccupation with sample sizes alone with disregard for other methodological considerations would yield a few very large, slow and expensive trials with a great potential for biases other than low statistical power. As noted by others, “unbiased trials with imprecise results trump no results at all” (Schulz & Grimes, 2005); and insisting on enlarging trials will simply shift the focus from one source of imprecision and noise (lack of statistical power) to another source of imprecision and noise (lack of methodological rigour and precision); only with fewer trials to show it. In all, improving the quality and efficiency of existing research may be more viable and rewarding than introducing completely new research designs, even with just adopting different practices for existing trials (Lakens, 2014). One opportunity is pilot testing, which will be discussed below.

† “When a distinguished but elderly scientist states that something is possible, they are almost certainly right. When they state that something is impossible, they are very probably wrong.” — Clarke’s first law.
The virtues of appropriate pilot testing

Maintaining the RCT design for the present, the previously mentioned effect of increasingly participant-hungry RCTs may be mitigated by using pilot studies; which may be external (i.e., not part of the main RCT) or internal (i.e., a part of the actual main RCT). In fact, the importance of pilot studies in Internet-based interventions cannot be overestimated given the challenges that were encountered during this trial (and other trials). A pilot study would cost some time to set up and run, but would deliver inestimable beneficial information for a main RCT in terms of data and experience on processes, resources, management and scientific parameters (Thabane et al., 2010). Specifically, of great value would be information on feasibility of actually performing a RCT, information on management and recruitment logistics, the likelihood of recruiting sufficient numbers of patients on time and acquiring estimates of variances, dropout numbers and other parameters of value to guide statistical estimations. Such pilots have also been used to, e.g., test the feasibility and cost-effectiveness of different patient recruitment strategies as a precursor to a full-fledged RCT (Jones, Goldsmith, Hewson, & Williams, 2013). In internal pilot testing, the application of planned interim analyses (i.e., statistical analyses on data collected so far) can be used to steer the main RCT as it is in progress.

Although unplanned interim analyses (or ‘data peeking’) can be abused to, e.g., stop recruiting participants just when statistical significance is reached (Pocock & Hughes, 1989; Stegert et al., 2015), planned interim analyses in the early phases of a RCT can serve as an anchoring point for population variance and sample size estimations. Rather than continuing the RCT (and ending up underpowered or overpowered), such an analysis can help to adjust the sample size, especially if there is a large discrepancy between the estimated sample size and the actual, preliminary sample size. This can be especially beneficial in trials requiring large sample sizes, such as non-inferiority or equivalence trials. Crucial, however, is that these analyses are pre-specified and preferably performed by statisticians blinded to treatment allocation (Friede & Kieser, 2003).
In the current Phobias Under Control trial, no formal pilot testing was performed. Firstly because of time constraints, but also under the assumption that previous research had provided us ample experience and evidence to successfully perform a large-scale multi-centre randomised controlled trial within the allotted time†. Though allowances must be made for researcher optimism and exuberance – scientific research would quickly go the way of the dodo without either – in retrospect the use of a small pilot or feasibility study would have considerably benefited our trial.

The vices of inappropriate pilot testing

In short, pilot studies can be a powerful tool to serve as a ‘test–run’ for a full RCT, streamlining data collection practices, serving to train staff and correct any issues that might have serious consequences in a full RCT. Unfortunately, many pilot studies probing Internet–based interventions have been inappropriately used for hypothesis testing or to generate an effect size (e.g., Mathiasen, Riper, Ehlers, Valentin, & Rosenberg, 2016; Robinson et al., 2014; Wallach, Safir, & Bar–Zvi, 2011). This is inappropriate for a number of reasons (Arain, Campbell, Cooper, & Lancaster, 2010). Firstly, although often used to generate a preliminary effect size to calculate a main RCT sample size – this idea makes sense, at least intuitively – the chances of such a pilot trial actually yielding a useful effect size are very small indeed (Kraemer, Mintz, Noda, Tinklenberg, & Yesavage, 2006).

One illustration of this issue, is that when expected effect sizes are medium or small, the pilot trial is invariably underpowered to detect such an effect size, leading to – at best – inconclusive results and an uninformative pilot trial; or perhaps the intervention is prematurely labelled ‘ineffective’ and abandoned completely. Secondly, and conversely, due to its inherent imprecision, a pilot trial may also lead to inflated effect sizes, misguiding eventual sample size estimations and leading to severely underpowered main RCTs. The most likely outcome of such a pilot trial is then, rather unhelpfully; a very imprecise estimation of an effect size (Kraemer et al., 2006). At best, pilot studies may be used to yield an estimate of the population variance (Teare et al., 2014).

† To borrow from Faulkner, in the end we were not pleased exactly, nor vindicated.
More pragmatically (and perhaps slightly more disappointingly), one should be willing to concede that failure is an option, and that the outcome of a pilot study can be that undertaking a full RCT would be inadvisable in terms of cost, logistics or sheer effort needed to recruit sufficient participants.

As stated earlier, however, any effect size begotten from a pilot study should not be the main driver for this decision. Despite its caveats and the lesser ‘prestige’ of pilot studies (Arain et al., 2010), the innovative nature of Internet–based interventions makes these interventions a prime candidate for – appropriate – pilot testing (Leon, Davis, & Kraemer, 2011), especially when reported clearly and transparently (Thabane et al., 2016). Such pilot trials, when published, can be an inestimable source of information for other researchers, disseminating crucial and relevant information quickly in the fast–paced field of Internet interventions.

**RESEARCH ON DROP–OUT & NONADHERENCE**

Non–adherence to psychotherapy occurs routinely in FtF psychotherapy (Fernandez et al., 2015), and although often used as an argument against using Internet–based interventions, it is by no means unique to these interventions. It has, however, been a focal point of research for well over a decade (Eysenbach, 2005) since it is readily measurable and patently evident in Internet–based interventions. Although a good deal of research effort has focussed on predictors of (Donkin et al., 2011; Karyotaki et al., 2015), and reasons for (Donkin & Glozier, 2012), drop–out from Internet–based interventions, a persistent confounder is that non–adherence to an intervention is often accompanied by drop–out from follow–up measurements (Cunningham, 2014). This was also found in the Phobias under Control intervention (Chapter 6), and has serious implications for the generalisability of information on non–adherence to the intervention; limiting predictors to baseline characteristics such as pre–treatment clinical scores and demographic variables such as age, sex and educational level. Due to the simple fact of not having much data from participants who are non–adherent to the intervention, little can be said on the reasons for these participants to stop using the intervention. Some may be trivial (e.g., computer problems) but others may be driven purely by the content of the intervention or the delivery.
While it has been argued that some may discontinue the intervention as they feel they have derived sufficient benefit from the Internet–based intervention (Hilvert–Bruce, Rossouw, Wong, Sunderland, & Andrews, 2012), this would seem implausible in the current case as all but four participants proceeded to FtF psychotherapy after the Internet–based intervention (Chapter 6). More qualitative and mixed–method research into the predictors and reasons for drop–out, as has been previously attempted (Bossen, Buskermolen, Veenhof, de Bakker, & Dekker, 2013; Donkin & Glozier, 2012; Postel, de Haan, ter Huurne, Becker, & de Jong, 2010; Varsi, Gammon, Wibe, & Ruland, 2013), would be welcome to further elucidate why participants choose to discontinue an intervention that is designed to help them. However, most research will, almost by definition, struggle with missing data from those who discontinued both the intervention and the follow–up measurements.

The definition & measurement of nonadherence

Analogous to non–Internet–based trials, adherence to other trials of interventions is described in a staggering variety of ways, with a recent review of medical trials finding no fewer than 44 different definitions of ‘adherence’ (Jeffery et al., 2014). This is no different in the field of Internet–based interventions (Kelders, Kok, Ossebaard, & Van Gemert–Pijnen, 2012), although some effort has been made to at least standardise the terminology of non–adherence, differentiating between not finishing the intervention itself, and not finishing follow–up measurements (Christensen et al., 2009). The measurement of non–adherence to Internet–based interventions is currently primitive: based on secondary observations, in essence it is based on whether particular content was displayed on a participant’s screen; with no regard for whether this content was actually read, understood, or acted upon (Chapter 6).

Unless specific feedback is received from the patient by way of homework exercises, the therapist has no idea whether the therapeutic content has been appropriated. As such, the current definitions and measurements of adherence to Internet–based interventions are, at best, vaguely defined and imprecise surrogate outcomes to a biomedical dose–response paradigm

† Vigorous debate on this issue continues to this day.
that may not even be applicable to the field of Internet–based interventions. This may seem a rather bold statement, but it does explain why small (or no) associations between adherence and treatment effect are being found (Donkin et al., 2011, 2013).

The decision of when a participant is (or is not) adherent is often made as a post–hoc assessment and based on arbitrary cut–off numbers. This is often done since there was no pre–defined cut–off score to start with (e.g., a participant should finish at least 4 out of 5 sessions to be marked ‘adherent’). Currently, post–hoc binning is often used to group participants in different categories (e.g., chapter 6) based on their adherence. In hindsight, this definition would have to be made beforehand, ideally pre–registered with the trial protocol, and defined in a standardised manner, using e.g. the ‘intended usage’ definition (Kelders et al., 2012), where clinicians reach consensus on which percentage or parts of an intervention should be completed to derive maximum benefit from the intervention. The transparency and comparability of adherence reporting in Internet–based trials would greatly improve from this.

Technical aspects

Sorely lacking then, is research not only in the content of the intervention (the message), but in its delivery format (the medium). Technology is never transparent, and any technology can be both a filter to treatment content (diminishing its therapeutic potential) and a catalyst to the treatment content, amplifying its therapeutic potential (chapter 7). Although all digital formats are ephemeral to some degree or other, it would not be impossible to research more general principles of Internet–based interventions that are not tied specifically to any kind of platform, device or software (i.e., technology–agnostic). One such example is research into the elements of design and psychology that encourage users to adhere to Internet–based interventions (Kelders et al., 2012), the results of which can be applied to any kind of delivery platform. Additionally, there is no reason to assume that ‘one size fits all’ in terms of design and delivery. However, there appears to be no research in adaptive design, much less in finding out which patients prefer which delivery format.
There is some discussion on whether the RCT is the optimal paradigm for research into Internet–based interventions, especially since these interventions seem to be happily implemented without any empirical substantiation whatsoever. Moreover, there is some evidence that RCTs in outpatient samples tend to result in unrepresentative samples (Stirman, DeRubeis, Crits–Christoph, & Brody, 2003), due to exclusion criteria, informed consent procedures and an unwillingness to be randomised. However, it is unclear whether the two examined proposed alternatives here are an acceptable and feasible alternative to RCTs, and whether the benefits (e.g., clinic–based randomisation) can outweigh the drawbacks (e.g., even larger sample sizes).

Rather than seeking for alternatives to RCTs or fixedly trying to increase sample sizes (leading to even longer patient recruitment periods), a lot can be gained from tackling the two most salient current issues: lack of precision; leading to uncertain or inconclusive results, and lack of speed; leading to research that lags behind clinical practice.

Firstly, the methodological quality and transparency in trial reporting can be improved by pre–registering all trials and by adhering to appropriate reporting standards such as CONSORT–eHealth (Eysenbach, 2011). Specifically, pre–specification of adherence definitions and measurements will lead to a marked improvement in the search for an association between adherence and treatment outcome.

Secondly, appropriate pilot testing could lead to more efficient, quicker RCTs by preventing major logistical issues such as patient non–adherence and drop–out.

Finally, making patient–level data available for research (such as individual patient–data analyses) would provide a wealth of information, especially for research questions that require very large sample sizes such as mediation or moderation analyses. However, very few journals prescribe routine data sharing at the moment.

Whatever ingenious alternative is found for RCTs, let us realise that there is no inherent reason why a RCT should be ‘slower’ to perform than
any other research design – the classical parallel–group design is actually relatively “quick” compared to quasi–experimental or cross–over designs. It is therefore not unthinkable that those disparaging RCTs for research on Internet–based interventions might in fact be bad workmen blaming their tools.

CONCLUDING REMARKS

The advent of very new technologies may once again turn the tables of therapist–supported Internet–based interventions. For example, the fledgling field of embodied conversational agents (Provoost, Lau, Ruwaard, & Riper, 2017) provides an interesting new addition to human–like, but non–human supported interventions. In short, these conversational agents are like computer–generated artificial therapists; intended to replace real–life guidance with human–like animations that provide guidance based on artificial intelligence†. Although currently in its infancy, it is not inconceivable that within the next 10 years some guided Internet–based interventions will be developed that offer credible support without actual human intervention. This would have a profound influence on mental health worldwide, since little or no human intervention is needed. If these agents can be disseminated to populations that have currently little or no access to psychological treatments, the reach and impact can be enormous. However, considerable technological, cultural and ethico–legal complications will need to be solved.

For phobias specifically, the field of Virtual Reality Exposure Therapy (VRET) has been considered ‘promising’ for decades, with research going to at least the mid–90s (Rothbaum et al., 1995). Its most salient benefit is that the therapist has total control over the exposure experience, placing the ‘dosage’ of exposure at the therapist’s fingertips. Moreover, it offers exposure to stimuli not readily obtained from an office, such as exotic animals; or situations which would be difficult or time–consuming to find (Emmelkamp, 2005). However, the technology has hitherto been prohibitively expensive,

† “Any sufficiently advanced technology is indistinguishable from magic.” — Clarke’s third law.
complex and cumbersome, making it impractical for reliably treating large numbers of patients in routine mental healthcare without, e.g., considerable down-time due to technical issues. Relatively low-cost, consumer-grade VR equipment has recently become available, both as computer-based hardware (e.g., Oculus Rift) and more recently, as add-on hardware to consumer-grade mobile phones (e.g., Samsung Gear VR).

Previous effectiveness research on VR has been positive (Opris et al., 2012; Parsons & Rizzo, 2008) and these effects seem to transfer from the virtual world to real-life (Morina, Ijntema, Meyerbröker, & Emmelkamp, 2015). However, the methodological quality of most of these trials is low; with none of the trials included in these meta-analytic reviews exceeding a sample size of 58, leading to a paltry 30 or fewer subjects per condition per study. Presumably, this is due to the same prohibitive factors such as cost that made VR impractical for clinics; but hopefully the increasing availability of affordable VR technology will soon spawn better quality trials, and some of these are already underway (Miloff et al., 2016).

The advent of smartphone-based interventions may also have a profound influence on how patients approach mental health, although this was not covered in this dissertation. Some years ago, a systematic review found no more than eight papers testing the efficacy of a smartphone-based intervention for mental health problems (Donker et al., 2013), and the evidence of other health-related apps was equally sparse (Edwards et al., 2016). Even though evidence of the effectiveness of these apps for mental health is mounting (Ly et al., 2015), there is a considerable uphill battle ahead with regards to disseminating these app-based interventions in a field populated with tens of thousands of apps, with no discerning characteristic to make the app known to the patient as being ‘evidence-based’.

FINAL THOUGHTS

Although we could not demonstrate cost-effectiveness of this intervention in the current population, further research is warranted to consider the fundamental requirements of successfully embedding the intervention in routine mental healthcare. Furthermore, with more research into patient preferences
and information on which patients actually benefit from Internet–based interventions, there is a great potential for these interventions to optimise mental healthcare delivery and quality. Examples include grand challenges such as increasing access to mental healthcare by using low–threshold interventions, lowering stigma by moving treatments away from psychologists’ offices, empowering patients by increasing self–efficacy and fitting in psychological treatments into a patient’s daily life.

This discussion underlined many caveats and limitation of not only the Phobias Under Control project, but also Internet–based interventions as a whole. However, awareness of limitations and challenges is a crucial prerequisite to improve these interventions and to lift the field to the next level.

As such, this discussion should not be seen as a conclusion — it is just the beginning.


Kenter, R. M. F., Cuijpers, P., Beekman, A., & van Straten, A. (2016). Effectiveness of a Web–based Guided Self–help Intervention for Outpatients With a Depressive Disorder: Short–term Results From a Randomized Controlled Trial. Journal of Medical Internet Research, 18(3), e80. https://doi.org/10.2196/jmir.4861


