BMJ Open

A cluster randomised controlled trial of a guided self-help mental health intervention in primary care

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-023481
Article Type:	Research
Date Submitted by the Author:	25-Apr-2018
Complete List of Authors:	Mathieson, Fiona; University of Otago – Wellington, Department of Psychological Medicine Stanley, James; University of Otago, Wellington, Public Health; University of Otago, Wellington Collings, Catherine (Sunny) ; University of Otago Wellington, Deans Department Tester, Rachel; University of Otago, Wellington, Primary Health Care and General Practice Dowell, Anthony; Wellington School of Medicine and Health Sciences, General Practice
Keywords:	MENTAL HEALTH, PRIMARY CARE, BRIEF INTERVENTIONS

A cluster randomised controlled trial of a guided self-help mental health intervention in primary care

Fiona Mathieson[1]
Email: fiona.mathieson@otago.ac.nz
James Stanley [2]
Email: james.stanley@otago.ac.nz
Sunny Collings[3]
Email: deanseauow@otago.ac.nz
Rachel Tester [4]
Email: rachel.tester@otago.ac.nz
Anthony Dowell[4]
Email: tony.dowell@otago.ac.nz

* Corresponding author: fiona.mathieson@otago.ac.nz Ph. 64 4 9186034, Fax 64 4 3855877
[1]Department of Psychological Medicine, University of Otago, PO Box 7343, Wellington South 6242, New Zealand
[2]Biostatistical Group, Dean's Department, University of Otago, PO Box 7343, Wellington South 6242, New Zealand
[3]Office of the Dean and Head of Campus Te Tari Manutaki,
Department of Psychological Medicine, University of Otago, PO Box 7343, Wellington South 6242, New Zealand
[4]Department of Primary Health Care and General Practice, University of Otago, PO Box 7343, Wellington South 6242, New Zealand

Word Count: 5801

Abstract

Objectives: To ascertain whether an ultra-brief intervention improves outcomes for patients in general practice with mild-to-moderate mental health concerns.

Trial design: Two-arm cluster randomised controlled trial.

Methods:

Participants: General practitioners (GPs) were invited based on working in a participating general practice. Patients were eligible to participate if aged 18-65, scored ≤35 on the Kessler-10 (K10) and if meeting local mental health access criteria (based on age, low income, or ethnic group).

Interventions: Intervention arm GPs were trained on the ultra-brief intervention (UBI) approach, with participating patients receiving three structured appointments over five weeks. GPs randomised to Practice as Usual (PAU) did not receive training, and delivered support following their existing practice approaches.

Outcome Measures: Primary outcome was patient-level K10 score at 6 months post-recruitment.

Randomisation: GP practices were randomised to UBI training or PAU at the start of the study.

Blinding: GPs were not blinded to group assignment.

Results:

Numbers randomised: 62 GPs (recruiting 85 patients) were randomised to UBI, and 50 to PAU (recruiting 75 patients).

Numbers analysed: 31 GPs recruited at least one patient in the UBI arm (70 patients analysed), and 21 GPs recruited at least one patient in the PAU arm (69 patients analysed).

Outcome: K10 scores from an intention-to-treat analysis were similar in UBI and PAU arms, with a wide confidence interval (mean adjusted K10 difference = 1.68 points higher in UBI arm, 95% CI -1.18, 4.55). Secondary outcomes were also similar in the two groups.

Conclusions: The UBI intervention did lead to better outcomes than practice as usual. Results from 'negative trials' such as this contribute to the continuing development of brief psychological therapy options for primary care.

Trial registration: Australia New Zealand Clinical Trials Registry ACTRN12613000041752

Funding: Compass Health, Oakley Mental Health Research Foundation, Wellington Medical Research Foundation, University of Otago Research Fund

Strengths and limitations

- Pragmatic effectiveness trial of a mental health intervention in primary care.
- Intervention included Maori cultural adaptations.
- Recruitment issues limit strength of results.
- Intervention was applied to more severe mental health presentations that it was developed for.
- GP degree of adherence to the intervention tool is unclear.

Keywords

Mental Health, Primary Care

Introduction

Mental health is major aspect of health and poor mental health is highly prevalent in the general community. Consistent with international findings, just under 40% of the New Zealand (NZ) population had met criteria for a diagnosable mental disorder during their life, and roughly a fifth had experienced a mental disorder in the previous year (Oakley-Browne, Wells, & Scott, 2006).

There is also considerable international concern about the healthcare burden arising from mental health problems and substance abuse (Greenberg et al., 2003; Horton, 2012; Wang, Simon, & Kessler, 2003)), with the World Mental Health Survey (of 21 countries) suggesting that only 41% of people with depression received treatment that met even minimal standards (Thornicroft et al., 2017).

In NZ, as in other OECD countries, mental health problems are common presentations in primary care. Around one-quarter of primary care patients (26.5% and 29.8% of men and women, respectively), attending their general practice in NZ met criteria for a mild-moderate mental health disorder and an estimated 50-70% of mental health concerns are managed exclusively at the primary care level, since secondary care services have become more targeted towards severe and enduring mental illness in recent years (The MaGPIe Research Group, 2003) .

Internationally there is a call for psychological therapies to be more widely available in primary care (Layard et al., 2006), and growing unease about increasing levels of antidepressant medications being prescribed compared with the limited resources available for psychological interventions (Hollinghurst, Kessler, Peters, & Gunnell, 2005). However, treatment options at the primary care level are limited, with GPs expressing concerns about gaps in services for patients with mild-moderate mental health presentations and a desire to offer a brief intervention themselves (Dowell et al., 2008; Garrett et al., 2007). In NZ, GPs reported that as few as 22% of patients with mild-moderate mental health syndromes receive any formal help (The MaGPIe Research Group, 2006).

Such patient presentations often comprise sub-threshold syndromes (American Psychiatric Association, 2013; Mathieson, Collings, & Dowell, 2009), and cases of mild-moderate common mental disorder. These are combinations of problems such as anxiety, depression, substance use and interpersonal problems that do not meet the threshold for disorder in

standard diagnostic systems such as DSM-5. Often these arise in the context of social problems or family or economic stress. In NZ, 36% of general practice attendees report anxiety, depression or substance-use, or a combination of these issues (The MaGPIe Research Group, 2003). Such presentations can be associated with significant impairment in functioning and suffering (Collings & MaGPIe Research Group, 2005; Wagner et al., 2000), with some going on to develop severe depression (R. Kessler, Zhao, Blazer, & Swartz, 1997; Sadek & Bona, 2000). Intervention may be warranted for up to 80% of those affected (The MaGPIe Research Group, 2006; Wagner et al., 2000), but referral out of the practice can be problematic due to referral eligibility criteria, waiting times, administrative issues and cost (Dew, Dowell, McLeod, Collings, & Bushnell, 2005; Dowell et al., 2007, 2008).

Increasing knowledge of the burden of mild-moderate disorder led to the development of a platform of Primary Mental Health Initiatives in NZ, which included some increase in access to psychological therapies and extended consultations with GPs. The inclusion criteria for these initiatives, however, mean that only up to 15% of the population can gain access to those services (Dowell et al., 2008).

This service-gap led us to develop a GP delivered ultra-brief intervention (UBI), with development and refinement based on service user feedback (Mathieson et al., 2013). This model has the advantages of avoiding the need for referral on to an expensive professional, such as a psychologist, of being easily accessible to patients, and of potentially building on existing trusted relationships. This fits with the movement towards alternative methods of service delivery for mild to moderate mental health presentations, often termed 'low intensity' interventions. These interventions often include guided self-help, bibliotherapy and computerised delivery of care, with current evidence suggesting that even minimal therapist contact leads to better outcomes than self-help alone (Gellatly et al., 2007; Gellatly et al., 2017; Jorm & Griffiths, 2006; D. Kessler et al., 2009).

UBI was feasibility tested with a group of 16 patients and then adapted for Maori (the indigenous people of New Zealand) and feasibility tested with a group of 9 patients (Collings et al., 2012; Mathieson, Mihaere, Collings, Dowell, & Stanley, 2012). Based on questionnaire feedback, clinician and patient satisfaction ratings for both feasibility studies were very positive in terms of relevance and acceptability. The psychological well-being of the patients, as measured by the Kessler-10 (K10) (R. Kessler et al., 2002), was also significantly improved post-intervention (at 3 month follow-up) for both Maori and non-Maori, although

there was no control group (Collings et al., 2012; Mathieson et al., 2012). Based on these initial findings we designed a cluster randomized controlled trial to measure the effectiveness of UBI.

The aims of the study were to compare patient-level outcomes on (1) mental health state (as measured by K10 scores) at 6 months between UBI and practice as usual (PAU) study arms (primary outcome) and (2) levels of distress (depression and anxiety) and functioning (work, social and relationship) at 8 weeks and 3 months between UBI and PAU study arms (as secondary outcomes).

Methods

A protocol for this study has been previously published, and includes description of planned analyses (Collings et al., 2015). The trial was registered prior to recruitment commencing with the Australia New Zealand Clinical Trials Registry (registration ACTRN12613000041752.)

Design

We used a pragmatic two-arm single blinded, cluster randomised controlled trial of UBI compared with PAU, in a primary care setting. GPs were randomised by practice to exclusively deliver either UBI or PAU to all their recruited patients. GPs were treated as the clusters in the study design (while there was be clustering by practice, the GPs were treated as the unit of analysis as practitioner attributes were anticipated to be a higher source of variability in outcomes.) Analysis followed an intention-to-treat approach.

Setting

The study was conducted in general practices in the greater Wellington region, New Zealand. This included practices in both city and semi-rural settings, serving populations from a wide range of socio-economic backgrounds. Recruitment took place between 1/5/2013 and 1/7/2016. The trial ended prior to achieving the final sample size when funding for data collection was exhausted.

Participants

This was a pragmatic trial supported within existing treatment services. GPs were eligible to participate if they were currently working in a practice that was part of the Compass Health Primary Health Organisation (PHO) which covers the greater Wellington region.

Patients were eligible if aged between 18 and 65 and identified by their GP in a routine appointment as experiencing stress or distress. To allow comparison of UBI with a PAU arm, all patients needed to meet access criteria of a local partner Primary Health Organisation (PHO) to psychological therapies. These groups were youth (defined as 18-24-years old), and individuals aged 25 years or older who were identified as low income, or Māori or Pacific Island ethnicity.

Patients were required to score 35 or less on the Kessler Psychological Distress Scale (K10) (Andrews & Slade, 2001; R. Kessler et al., 2002) during their initial GP consultation, with no lower cut-off on this score. The present study followed previous study protocols (Collings et al., 2012; Mathieson et al., 2012) by including scores between 30 and 35 on the K10 as indicative of mild to moderate levels of psychological distress rather than major psychiatric disorder. Individuals taking anti-depressant or other psychiatric medications were eligible to participate in the study.

Patients were excluded if they lacked fluency in English (as the intervention is an English-language based 'talking therapy'); had significant levels of cognitive impairment as determined by the GP; or had reported recent or acute suicidal ideation (i.e., within the previous 2 weeks). Chronic low level suicidality did not exclude an individual from participating. However, GPs were informed of patients who had high scores or suicidality at screening, or for whom referral to appropriate (secondary) mental health services by GPs was indicated, and these patients were not eligible to participate further in the study.

Inclusion criteria were based on the access criteria of a local partner primary health organisation (PHO) to psychological therapies. These criteria were youth (defined as 18-24 years old), or individuals aged 25 years or older with low income, or Māori or Pacific Island heritage.

Recruitment of practices and GPs

Initial recruitment of practices was supported by the partner PHO. GPs were identified using primary health organisation and practice lists. All of the practices contracted under the partner PHO were contacted (N=52) and invited to participate in the study, and an effort was made to contact all of the GPs within these practices by email, telephone or in person. A total of 23 practices initially consented to participate in the study and a further 18 were recruited during the course of the study. Two practices merged and three withdrew (in each case the single participating GP left the practice) leaving a total of 37 practices involved in the study.

Randomisation of practices to study arms

Consenting practices were randomised to provide either UBI or PAU to eligible patients. Randomisation was conducted at the practice level to reduce the risk of contamination if GPs from the same practice were assigned to opposite study arms. To ensure approximately equal numbers of GPs per study arm, randomisation of practices was conducted within five strata, according to the number of participating GPs (one/two/three/four/more than four). An additional two practices dedicated to youth health that were not part of the partner PHO were included and randomised into each arm of the study (i.e. these two practices formed their own stratum). Randomisation was performed following individual GP consent as a single step, with randomisation conducted by the project biostatistician (JS).

GPs randomised to the UBI study arm completed a single two-hour training session (as previously described (Collings et al., 2012)). Due to the training nature of the intervention, it was not possible to blind GPs as to their study arm allocation.

Recruitment procedures

GPs identified patients with common mental health problems who might fulfil study criteria during routine appointments. These patients were screened by the GP for eligibility (using the K10), and referred to the study team. A research assistant then contacted potential participating patients, met with them in person where possible to explain the study, confirm eligibility, obtain consent to participate, and collect pre-treatment (baseline) data. Measures were then collected by mail or email at post-treatment (8 weeks, 3 months and 6 months). Patients received compensation (NZ $30 [US$21] vouchers, and entry into a draw for an

iPad) following the completion of the final questionnaire, to recompense for time and effort in participating in the study.

Intervention

UBI is guided self-management programme which can be delivered by a GP after a single two-hour training session using a treatment manual based on structured problem solving, motivational interviewing and cognitive behaviour therapy (supported with self-help booklets on relationships, bodily stress, breaking habits and stress management).

Patients who consented and completed the intake data collection (K10 and baseline measurements) received the GP-led intervention in three short, structured face-to-face sessions (one 30 and two 15 minute sessions) over a five to six week period. Relevant booklets were provided to the patient after the first session, to be used in the following session. The study protocol allowed for patients in either study arm to alter their treatment as needed (e.g. access other talking therapies, or commence mental health medications). Patients were blinded as to their study allocation in that patients in PAU practices were not informed that the UBI was offered in practices randomised to deliver UBI. They were simply told that the study was looking at the effectiveness of PAU.

Practice as usual

Patients in the PAU study arm received GP support delivered according to their practice as usual (and available existing services). PAU typically consists of supportive counselling in a 15 minute face-to-face consultation, the provision of psychotropic medication, referral to psychological or other counselling options, or referral to relevant community services.

Patient characteristics

Patients are described on the basis of age, gender, prioritised ethnicity and NZiDep, a NZ-developed index (Ministry of Health, 2004) of individual-level socioeconomic deprivation. GPs in practices assigned to the PAU study arm received optional training in the intervention at the end of the study.

Patient and Public Involvement

This study had an academic mental health consumer as part of the research team at the feasibility stage, and designed the intervention based on feedback from a focus group process with potential patient users of the mental health intervention which asked what characteristics such an intervention would need to have. This collaborative process is fully described in (Mathieson et al., 2013). This RCT did not have academic consumer or patient involvement in the recruitment to and conduct of the study and the burden of the intervention was not assessed by the patients. Results of this study will be disseminated by email to GP participants who indicated they wanted them on the consent form.

Outcome measures

The primary outcome measure was the K10 scale (Andrews & Slade, 2001; R. Kessler et al., 2002) score at 6 months (adjusted for score at baseline: see analysis). The K10 is widely used as a clinical outcome measure in Primary Care and General Practice in NZ (Dowell et al., 2008). All analyses were conducted to look at patient-level outcomes.

Secondary outcomes were:

1) Hospital Anxiety and Depression Scale (HADS), which measures the severity of depressive and anxiety symptoms in outpatient hospital settings (Zigmond & Snaith, 1983). Reductions in HADS score indicate reduced anxiety and depression.
2) Comparison of K10 scores by treatment group at 8 weeks and 12 weeks, adjusted for baseline scores (to capture short and medium term effectiveness).
3) Work and Social Adjustment Scale (Mundt, Marks, Shear, & Greist, 2002), a measure of work, social and relationship functioning) administered at baseline, 8, 12 and 26 weeks.

Outcomes were measured at the same time points in both UBI and PAU groups (baseline, and at 8, 12, and 26 weeks following baseline)

Statistical methods

Sample size and Power analysis

Sample size for the cluster randomised trial was calculated using a simulation method, using standard deviations of patient outcomes from the UBI feasibility study (standard deviation of post-treatment scores = 7.5; unpublished data). To detect a difference in K10 improvement

scores of 6 points in the UBI arm compared with 2 points in the control arm (at 80% power and alpha = 0.05) would require 15 GPs per arm recruiting eight completing patients each on average (n=240 total with complete data). Adjusting for loss to follow-up of 20% gave a recruitment target of ten patients per GP. The simulation settings roughly correspond to an intraclass correlation (ICC) of 0.15 for considering clustering of patient scores by GP (equivalent to the ICC from the feasibility study; unpublished data). Power analysis for the secondary HADS outcome indicated 80% power to detect a difference of 3.2 points between groups (based on a standard deviation of approximately 6 (Spinhoven et al., 1997)) assuming a similar ICC for the HADS scale as for the K10 measure (empirical data were not available).

Data Analysis

The statistician was blinded to the intervention or control status of participants (both practices and patients) during conduct of the study and analysis. Results were unblinded once analysis was complete. Data processing and analysis were conducted in R 3.2.3 (R Institute, Vienna) with linear mixed models fit using the lmer package (Bates, Maechler, Bolker, & Walker, 2015).

For the primary outcome, K10 scores at 6 months were compared between the intervention and control groups using mixed linear models (comparing post-intervention scores between groups, adjusting for intake score as a covariate, and treating GP clusters as random slope effects). Analysis was conducted on an intention-to-treat basis according to the study arm for each patient at entry into the study. Analyses were adjusted for all other baseline covariates (age, gender, ethnicity, educational level, and NZiDep).

Missing data were handled through the mixed linear models approach to the data, which allows for patients with missing data on the final outcome to be included in analyses, which in effect estimates a final outcome value conditional on the observed data at other follow-up times (i.e. validity being predicated under the assumption that the missing observations are missing at random, conditional on the observed data (Beunckens, Molenberghs, & Kenward, 2005; DeSouza, Legedza, & Sankoh, 2009)). The null hypothesis for this test was that the K10 scores at 26 weeks (adjusted for baseline score) were not different for the intervention and control groups.

For the secondary analysis, differences in mean scores on the K10 outcome were reported at 8 weeks and 3 months (using the same methods as above, within the mixed linear models framework). Analysis of the HADS and WSAS scores at 8 weeks, 3 months and 6 months utilised the same methods as for the K10 outcome.

Intra-class correlation coefficient (ICC) values were calculated for each outcome measure as a summary of clustering according to GPs. Details of the calculation methods are provided in the Supplementary Materials.

Additional treatments received during the trial (including medication and talking therapies) were analysed by study arm, based on self-report data collected at the 6 month follow-up. This descriptive analysis was not specified in the study protocol.

Confidentiality and data management

Consenting patients had their rights explained along with provision for data confidentiality. Paper and digital copies of the data were secured in locked storage on the premises of the University of Otago, Wellington. The questionnaire data was de-identified and entered into a spreadsheet for subsequent analysis.

Ethics approval

Ethical approval was received from the Health and Disability Ethics Committees (HDEC), Ministry of Health (Northern B Health and Disability ethics committee 12/NTB/2).

Adverse events were not anticipated in this trial, and arrangements were made to feedback clinical information to GPs if deemed necessary (e.g., high K10 scores or concerning self-reported statements about a patient's safety) in the course of data collection.

Results

GP Participants

A total of 41 practices agreed to participate, with a total of 112 individual GPs consenting to take part in the study (n=62 for UBI, and n=50 for PAU). Of these GPs, 31 recruited at least

one patient into the study in the UBI arm, and 21 recruited at least one patient in the PAU arm (see Supplementary Table R1).

Patient Participants

Figure 1 summarises the flow of patients into the study and participation in the interventions and follow-up. A total of 198 patients were referred into the study, and 160 met eligibility criteria and completed baseline assessments. The vast majority of these completed at least one post-intervention follow-up (70 / 85 in the UBI arm [82%]; and 69/75 in the PAU group [92%]) and hence contributed to the data analysis.

<Insert figure 1 about here>

Baseline data

Baseline sociodemographic characteristics of patients are presented in Table 1 for the two study arms. The two groups were roughly comparable at baseline, with a few more male participants and a slightly younger age profile in the UBI arm, but with a greater representation of females in the study overall.

Table 1. Patient sociodemographic profile by study arm.

Variable	Level	Study Group	
		UBI (Total n = 85)	PAU (Total n = 75)
		n (%)	n (%)
Gender			
	Female*	56 (65.9)	57 (76.0)
	Male	29 (34.1)	18 (24.0)
Age Group			
	15-24	55 (64.7)	37 (49.3)
	25-34	16 (18.8)	15 (20.0)
	35-44	3 (3.5)	13 (17.3)
	45-54	5 (5.9)	6 (8.0)
	55+	6 (7.1)	4 (5.3)
Ethnicity			
	NZE/Other	61 (71.8)	54 (72.0)
	Māori	19 (22.4)	14 (18.7)
	Pacific	4 (4.7)	2 (2.7)
	Asian	1 (1.2)	5 (6.7)
Highest education			
	At least secondary	78 (91.8)	71 (94.7)
	No secondary level	7 (8.2)	4 (5.3)
NZiDep			

0 (least deprived)	18 (21.2)	11 (14.7)
1	16 (18.8)	17 (22.7)
2	15 (17.6)	11 (14.7)
3	10 (11.8)	10 (13.3)
4	9 (10.6)	12 (16.0)
5 (most deprived)	17 (20.0)	14 (18.7)

* Includes one individual self-identifying as Female (transgender)

Mean baseline scores on the outcome measures were also similar between the two groups (Table 2, showing means and standard deviations). Boxplots of the distribution of baseline scores on each outcome scale are given in Supplementary Figure R1.

Table 2. Mean (standard deviation) of baseline scores for outcome measures by study arm

Outcome variable	Study Group	
	UBI (Total n = 85)	PAU (Total n = 75)
	mean (sd)	mean (sd)
K10*	29.5 (6.2)	28.1 (5.7)
HADS – total	20.6 (5.9)	19.5 (5.1)
HADS – anxiety	12.1 (3.6)	11.9 (3.5)
HADS – depression	8.5 (3.5)	7.7 (3.6)
WSAS	23.0 (8.2)	19.6 (8.5)
Health Thermometer**	55.4 (19.9)	58.8 (18.7)

* One patient in PAU group missing baseline value.
** Higher scores on the health thermometer indicate better health.

Health Outcomes at Follow-up

For the K10 primary outcome at 6 months the mean difference for UBI compared to PAU arm favoured the PAU arm (mean difference = 1.68, 95% CI 1.18, 4.55; p = 0.255), as shown in Table 3 (where positive differences indicate a better outcome for the PAU than UBI arm) While this result indicated no significant difference in K10 scores between the UBI and PAU arms (see Figure 2), each group had a reasonable improvement in K10 score from baseline (see Supplementary Table R1: for the PAU group mean improvement = 7.6, 95% CI 5.5, 9.6; and for the UBI group mean improvement = 5.9, 95% CI 4.0, 7.8).

Table 3. Mean difference in primary and secondary outcomes

	Mean difference (UBI minus PAU)*					
	8 weeks		3 months		6 months	
Outcome variable	mean diff (95% CI)	p	mean diff (95% CI)	p	mean diff (95% CI)	p
Primary outcomes**						
K10	-0.19 (-2.55, 2.16)	0.872	1.53 (-0.79, 3.84)	0.203	1.68 (-1.18, 4.55)	0.255
HADS	0.57 (-1.68, 2.82)	0.620	0.86 (-1.38, 3.10)	0.456	1.85 (-0.62, 4.31)	0.149
Secondary outcomes**						
HADS-A	0.27 (-1.02, 1.56)	0.684	0.70 (-0.60, 2.00)	0.296	1.05 (-0.39, 2.50)	0.161
HADS-D	0.39 (-0.82, 1.60)	0.533	0.24 (-0.96, 1.44)	0.701	0.88 (-0.38, 2.14)	0.178
WSAS	0.49 (-2.40, 3.38)	0.740	1.32 (-1.58, 4.22)	0.377	0.45 (-2.47, 3.37)	0.762
Health Thermometer	2.84 (-3.64, 9.31)	0.395	1.90 (-4.59, 8.39)	0.569	4.93 (-1.77, 11.62)	0.156

* Positive differences indicate better improvement in PAU than UBI arm.

** Number of participants contributing data to each analysis: UBI n = 70, PAU n = 69 (except for K10: PAU n = 68)

<insert Figure 2 about here>

The mean difference on the HADS measure at 6 months between UBI and PAU measures was 1.85 (95% CI = -0.62, 4.31, p = 0.149; see Table 3), though both groups again showed an improvement in mean score from baseline (Supplementary Table R1). Mean scores at each follow-up time are presented in Figure 3.

<insert Figure 3 about here>

Similarly, for all secondary outcome measures (HADS Anxiety and Depression sub-scales, WSAS, and Health Thermometer), the difference in outcomes at 6 months showed no significant advantage for either UBI or PAU measures (with relatively broad confidence intervals for these differences: see Table 3.)

Estimates of secondary analyses of outcomes at earlier follow-up times (8 weeks and 3 months) are also presented in Table 3. Differences between UBI and PAU were generally most pronounced at the final follow-up (6 months) compared to the interim follow-ups. Trajectories for mean scores in each group are presented in Supplementary Figure R2, Supplementary Figure R3, Supplementary Figure R4 and Supplementary Figure R5.

Ancillary analyses

Information on types of additional treatment received is presented in Supplementary Table R3, for those who completed the 6-month follow-up assessment (summary not specified in protocol). Similar proportions of completing patients between study arms were either on medication for mental health condition(s) at the beginning on the trial (UBI = 31%; PAU 25%), or started medication during the trial (UBI=18%; PAU=25%). Access to extended GP consultations or counselling sessions was higher for the PAU arm than for UBI (no UBI patient had an extended GP consultation, compared to 29% of PAU patients; and 25% of UBI patients had one or more counselling sessions, compared to 64% of PAU patients.)

Intra-class correlation coefficients (ICCs) for the outcome measures are presented in Supplementary Table R4. For the K10 (ICC = 0.129, 95% CI 0.045 – 0.231) this was relatively close to the ICC values used in planning the sample size for the study.

Discussion

The brief psychological treatment (UBI) delivered by GPs in New Zealand in routine practice settings did not lead to better outcomes than PAU in this pragmatic efficacy trial.

UBI was not significantly more effective than PAU in reducing distress as measured by the K10. The K10 was originally introduced as an assessment measure of psychological distress, but has also been used to track change in mental health status following intervention (Sunderland, Wong, Hilvert-Bruce, & Andrews, 2012). There were no significant differences in the secondary measures either.

We were unable to achieve full recruitment to match the pre-determined sample size. As such, we were unable to rule out non-inferiority of the intervention (UBI) compared to PAU in reducing the disability and distress associated with mild to moderate mental health problems: the bounds of the confidence intervals for the two main outcomes (K10 and HADS measures) included sizable-magnitude better outcomes for PAU over UBI (e.g. the upper bound for the K10 was a 4.55 point advantage for PAU).

Both UBI and PAU arms showed improvement in clinical outcome over the 6 month course of the study. These findings are in keeping with other work which demonstrates clinical effectiveness of brief psychological interventions in primary care settings (Cape et al., 2010).

These results suggest that GPs in both arms were achieving clinical benefit. We cannot rule out that UBI performs slightly worse than PAU, but our results are inconclusive due to our reduced sample size.

Strengths of this study

We consider the results of this trial a useful addition to the literature for two reasons. Firstly they describe the introduction of potentially useful adjuncts to existing therapy approaches in primary care in a randomised controlled setting, and secondly the 'negative results' raise questions about the challenges of conducting pragmatic trials of psychological interventions in primary care and also about the nature and effectiveness of PAU treatments. Feedback received from GPs during the training sessions suggested that elements of the UBI such as active listening, goal-setting; making a specific plan and following up on it are already used in routine practice. UBI had previously been piloted and shown to be both feasible and

acceptable to both clinicians and patients in a general practice setting (Collings et al., 2012). It was also able to be adapted in a culturally responsive way (Mathieson et al., 2012). During the course of the trial and following its completion there has been significant interest expressed by both patients and GPs in obtaining copies of the booklets and using elements of the UBI approach in routine consultations. Verbal feedback suggests that GPs particularly liked the helpful/unhelpful behaviour chart which was used to discuss how problems were maintained, the explicit linking of emotional responses to physical symptoms and the use of commitment and capability rulers (a motivational interviewing strategy).

There is an active debate about the optimal balance of intervention components for the management of common mental health problems, with an increasingly varied range of options available. Patients potentially have access to traditional face to face intervention with a therapist, access to materials available on the internet, and further access to rapidly developing telemedicine and virtual consultation options (Andersson, Carlbring, & Hadjistavropoulos, 2017; Gilbody et al., 2015). Our study shows that over the course of the trial, patients and GPs were able to adapt the standard pattern of the GP consultation to a series of three sessions, allowing a more participation from the patient. This ability to 'disrupt' the traditional pattern of GP consultations is important in an era where there is recognition in New Zealand and other OECD countries about the need to respond to the changing context of primary care, particularly in relation to long term conditions including common mental health problems (Baird et al., 2014).

Limitations

The difficulties in recruiting a sufficient sample size meant we were unable to establish benefit or rule out substantial inferiority of UBI compared to PAU. The main challenges of recruitment for trials in mental health have been described (Mason et al., 2007; McDonald et al., 2006; Weisfeld, English, & Claiborne, 2012). The current study contained specific additional challenges as outlined below.

Firstly, our recruitment was limited by specific entry criteria required by a funder (to allow access to treatments as part of the PAU group). This meant we did not meet our sample size target despite energetic problem-solving over a 3 year recruitment period. It also meant that many GPs were not using the UBI tool until weeks or even months after training. This casts

doubt on how well GPs would have adhered to the approach or recalled the principles, potentially affecting the quality of the intervention delivered.

Secondly, in this New Zealand context, the GPs in the PAU group had access to a sophisticated range of therapy options which included providing extended consultations themselves, as well as referring patients to psychological therapies such as counselling or CBT delivered by clinical psychologists (Dowell 2009). In addition, during the course of the study there were significant changes to the way in which the external psychological services were delivered in our local PHO, with therapists (mental health professionals) being placed within practices rather than at a central location making it easier for in-house referral. Thus the results may not generalise to settings where these additional therapies are unavailable in day-to-day practice.

These changes made the task of demonstrating non-inferiority more challenging. UBI is consistent with the contemporary primary care stepped care approach that tailors interventions to symptom severity and response to treatment (Dowell, Morris, Dodds, & McLoughlin, 2012). The intervention tool (UBI) used in this study was developed for sub-threshold mental health syndromes, but was, in practice, applied to moderate-to-severe problems, due to demand from GPs who said they needed higher thresholds in order to be able to recruit patients. In the New Zealand context it appears those needing mental health interventions in primary care have more severe problems than the tool was intended for. The intervention may have performed relatively better than PAU if applied to a mild-to-moderate group, but this would need further research to ascertain. The moderate-to-severe group are likely to require longer, more intensive interventions for it to make a difference.

Given the known efficacy of the PAU intervention in this setting (Dowell et al., 2008), the results also attest to the success of the PAU options rather than a specific failing of the intervention. Clinicians who participated in this study might be expected to be those who were motivated and skilled in supporting patients with mental health problems. It is unclear in this case the extent to which the GPs in the UBI treatment arm were adhering to the structured approach outlined in the treatment manual. Fidelity and adherence to training for psychological intervention has been subject to commentary in the literature (Bellg et al., 2004; Morton et al., 2016) and it is unclear as to the extent to which UBI GPs were able to adhere to the structured manual.

Conclusion

In this study both the PAU and UBI groups showed improvement in clinical outcome, despite UBI failing to demonstrate superiority or conclusive non-inferiority compared to PAU. This leaves open the question of whether this style of intervention may have potential value in a primary care setting, or whether some elements of this style of intervention are already being applied in practice by some clinicians. Either way, our results did not show that the UBI added value to usual care with patients with moderate-to-severe symptoms.

An ultra-brief approach such as UBI may add value if restricted to patients with mild mental health problems, as part of a suite of options, with different levels of intensity available to GPs in the primary care setting.

There is a significant need for further research into these issues, given the recognition of mental health problems at a community level (The MaGPIe Research Group, 2003; Whiteford et al., 2013) and the challenge of providing access to psychological therapy in an effective and cost-effective way (Clark, 2011; Gyani, Shafran, Layard, & Clark, 2013).

Figure Legends

Figure 1. Study flowchart of patient participation.

Figure 2. Mean K10 score (95% CI) at baseline and follow up for UBI and PAU study arms.

Figure 3. Mean total HADS score (95% CI) at baseline and follow up for UBI and PAU study arms.

Abbreviations

UBI: Ultra-brief intervention; PAU: Practice as Usual; GP: General Practitioner; PHO: Primary Health Organisation; K10: Kessler Psychological Distress Scale; HADS: Hospital Anxiety and Depression Scale; WSAS: Work and Social Adjustment Scale; NZDep2006: New Zealand index of individual socioeconomic deprivation.

Authors' contributions

All authors contributed to the study design and study protocol. FM and SC are co-principal investigators. SC conceived the study, obtained initial funding, and contributed to the development of the intervention. FM and RT obtained co-funding. FM largely developed the

intervention, led GP training and PHO liaison. AD contributed to the intervention design and GP training. JS contributed to the study design and designed and conducted the analysis. JS, FM, AD and SC jointly interpreted the results. RT contributed as research assistant, assisted with practice recruitment and GP training, led the patient recruitment, data collection, processing and project management in the latter stages. All authors contributed to and approved the final manuscript.

Funding

Funding for this study was received from the local partner PHO, Compass Health, The Oakley Mental Health Research Foundation, Wellington Medical Research Foundation, and the University of Otago Research Fund. Funding was also received from the Health Research Council of New Zealand for the initial feasibility study, and from Oakley Mental Health Research Foundation and the University of Otago Research Fund for the Māori adaptation. The funding bodies had no role in the study design, collection, analysis and interpretation of data or in the writing of the manuscript.

Competing Interests

None

Acknowledgements

Thanks to Gabrielle Jenkin, who managed the project in the early phases; Felicity Goodyear Smith, Simon Hatcher and Sarah Gordon, who contributed to the study design in the early phases; Amy Munstermann who facilitated liaison with Compass Health; Brigitte Lane, who recruited practices and led the data collection for the first year and Denise Steers who contributed as a research assistant.

Data sharing

Individual-level patient data are not available to other researchers as participants were not asked for consent to share their data. The study protocol (including statistical analysis plan) is available at in (Collings et al., 2015) (DOI:10.1186/s13063-015-0778-y). The code used to conduct the statistical analysis is available from the second author on request (james.stanley@otago.ac.nz).

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Washington, DC.

Andersson, G., Carlbring, P., & Hadjistavropoulos, H. D. (2017). Internet-based cognitive behavior therapy *The science of cognitive behavioral therapy* (pp. 531-549). San Diego, CA: Elsevier Academic Press; US.

Andrews, G., & Slade, T. (2001). Interpreting scores on the Kessler Psychological Distress Scale (K10). *Australian and New Zealand Journal of Public Health, 25*(6), 494-497.

Baird, M., Blount, A., Brungardt, S., Dickinson, P., Dietrich, A., Epperly, T., . . . McDaniel, S. (2014). Joint principles: integrating behavioral health care into the patient-centered medical home. *The Annals of Family Medicine, 12*(2), 183-185.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi: 10.18637/jss.v067.i01

Bellg, A. J., Borrelli, B., B., R., Hecht, J., Minicucci, D. S., Ory, M., . . . Czajkowski, S. (2004). Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology, 23*(5), 443-451.

Beunckens, C., Molenberghs, G., & Kenward, M. G. (2005). Direct likelihood analysis versus simple forms of imputation for missing data in randomized clinical trials. *Clinical Trials, 2*(5), 379-386.

Cape, J., Whittington, C., Buszewicz, M., Wallace, P., Underwood, L., & Medicine, B. (2010). Brief psychological therapies for anxiety and depression in primary care: meta-analysis and meta-regression. *8*(38).

Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: the IAPT experience. *International review of psychiatry, Aug 1;23*(4), 318-327.

Collings, S., & MaGPIe Research Group. (2005). Disability and the detection of mental disorder in primary care. *Social Psychiatry & Psychiatric Epidemiology, 40*(12), 994-1002.

Collings, S., Mathieson, F., Dowell, A., Stanley, J., Hatcher, S., Goodyear-Smith, F., . . . Munsterman, A. (2015). Clinical effectiveness of an ultra-brief intervention for common mental health syndromes in primary care: study protocol for a cluster randomized controlled trial. *Trials [Electronic Resource], 16*(260), 1-7. doi: 10.1186/s13063-015-0778-y

Collings, S., Mathieson, F., Dowell, A., Stanley, J., Jenkin, G., Goodyear-Smith, F., & Hatcher, S. (2012). Acceptability of a guided self-help mental health intervention in general practice. [Research Support, Non-U.S. Gov't]. *Family Practice, 29*(1), 43-49.

DeSouza, C. M., Legedza, A. T., & Sankoh, A. J. (2009). An overview of practical approaches for handling missing data in clinical trials. *Journal of Biopharmaceutical Statistics, 19*(6), 1055-1073. doi: doi: 10.1080/10543400903242795

Dew, K., Dowell, A., McLeod, D., Collings, S., & Bushnell, J. (2005). "This glorious twilight zone of uncertainty": mental health consultations in general practice in New Zealand. *Social Science & Medicine 61*(6), 1189-1200.

Dowell, A., Garrett, S., Collings, S., McBain, L., McKinlay, E., & Stanley, J. (2007). Primary Mental Health Initiatives: Interim Report (pp. 242). Wellington: School of Medicine & Health Sciences, University of Otago, Wellington.

Dowell, A., Garrett, S., Collings, S., McBain, L., McKinlay, E., & Stanley, J. (2008). Evaluation of the Primary Mental Health Initiatives: summary report 2008. Wellington: Otago University Wellington and Ministry of Health.

Dowell, A., Morris, C., Dodds, T., & McLoughlin, B. (2012). Psychological interventions in primary care mental health. In I. G. (Ed.), *Companion to Primary Care Mental Health*. London: Radcliffe Publishing.

Garrett, S., Dowell, A., Collings, S., McKinlay, E., McBain, L., & Stanley, J. (2007). Evaluation of the Primary Mental Health Pilot Initiatives (interim report): Unpublished report, Wellington.

Gellatly, J., Bower, P., Hennessy, S., Richards, D., Gilbody, S., & Lovell, K. (2007). What makes self-help interventions effective in the management of depressive symptoms? Meta-analysis and meta-regression. *Psychological Medicine, 37*(9), 1217-1228.

Gellatly, J., Chisnall, L., Seccombe, N., Ragan, K., Lidbetter, N., & Cavanagh, K. (2017). @Home eTherapy Service for People with Common Mental Health Problems: An Evaluation. *Behavioural and Cognitive Psychotherapy*, 1-6. doi: doi:10.1017/S1352465817000297

Gilbody, S., Littlewood, E., Hewitt, C., Brierley, G., Tharmanathan, P., Araya, R., . . . D., K. (2015). Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. *British Medical Journal, 351*(h5627).

Greenberg, P. E., Kessler, R. C., Birnbaum, H. G., Leong, S. A., Lowe, S. W., Berglund, P. A., & Corey-Lisle, P. K. (2003). The economic burden of depression in the United States: How did it change between 1990 and 2000? [Empirical Study; Quantitative Study]. *The Journal of Clinical Psychiatry, 64*(12), 1465-1475. doi: http://dx.doi.org/10.4088/JCP.v64n1211

Gyani, A., Shafran, R., Layard, R., & Clark, D. M. (2013). Enhancing recovery rates: lessons from year one of IAPT. *Behaviour Research and Therapy, Sep 30;51*(9), 597-606.

Hollinghurst, S., Kessler, D., Peters, T. J., & Gunnell, D. (2005). Opportunity cost of antidepressant prescribing in England: Analysis of routine data. *BMJ: British Medical Journal, 330*(7498), 999-1000. doi: http://dx.doi.org/10.1136/bmj.38377.715799.F7

Horton, R. (2012). Understanding disease, injury and risk. *The Lancet, 380*, 2053-2054.

Jorm, A. F., & Griffiths, K. M. (2006). Population promotion of informal self help strategies for early intervention against depression and anxiety. *Psychological Medicine, 36*, 3-6.

Kessler, D., Lewis, G., Kaur, S., Wiles, N., King, M., Weich, S., . . . Peters, T. J. (2009). Therapist-delivered internet psychotherapy for depression in primary care: A randomised controlled trial. [Empirical Study; Quantitative Study]. *The Lancet, 374*(9690), 628-634. doi: http://dx.doi.org/10.1016/S0140-6736%2809%2961257-5

Kessler, R., Andrews, G., Clope, L. J., Hiripi, E., Mroczek, D. K., Normand, S., . . . Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine, 32*, 959-976.

Kessler, R., Zhao, S., Blazer, D., & Swartz, M. (1997). Prevalence, correlates, and course of minor depression in the national comorbidity survey. *Journal of Affective Disorders, 45*(1-2), 19-30.

Layard, R., Clark, D., Bell, S., Knapp, M., Meacher, B., Priebe, S., . . . Wright, B. (2006). The depression report; A new deal for depression and anxiety disorders. In T. C. f. E. P. s. M. H. P. Group (Ed.): London School of Economics.

Mason, V. L., Shaw, A., Wiles, N. J., Mulligan, J., Peters, T. J., Sharp, D., & Lewis, G. (2007). GPs' experiences of primary care mental health research: a qualitative study of the barriers to recruitment. *Family Practice, 24*(5), 518-525.

Mathieson, F., Collings, S., & Dowell, A. (2009). Sub-threshold mental health syndromes: Finding an alternative to the medication of unhappiness. *Journal of Primary Health Care, 1*(1), 74-77.

Mathieson, F., Collings, S., Dowell, A., Goodyear-Smith, F., Stanley, J., & Hatcher, S. (2013). Collaborative research: A case example of dissemination of CBT in primary care. [Empirical Study; Interview; Focus Group; Qualitative Study]. *the Cognitive Behaviour Therapist Vol 6 2013, ArtID e4, 6*.

Mathieson, F., Mihaere, K., Collings, S., Dowell, A., & Stanley, J. (2012). Maori cultural adaptation of a brief mental health intervention in primary care. *Journal of Primary Health Care, 4*(3), 231-238.

McDonald, A. M., Knight, R. C., Campbell, M. K., Entwistle, V. A., Grant, A. M., Cook, J. A., . . . Snowdon, C. (2006). What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. *Trials, 7*(7), 1-9.

Ministry of Health. (2004). *Ethnicity Data Protocols for the Health and Disability Sector*. Wellington, New Zealand.

Morton, L., Elliott, A., Thomas, R., Cleland, J., Deary, V., & Burton, C. (2016). Developmental study of treatment fidelity, safety and acceptability of a Symptoms Clinic intervention delivered by General Practitioners to patients with multiple medically unexplained symptoms. *Journal of psychosomatic research., May 31*(84), 37-43.

Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. H. (2002). The work and social adjustment scale: a simple measure of impairment in functioning. *British Journal of Psychiatry, 180*, 461-464.

Oakley-Browne, M., Wells, J., & Scott, K. (2006). Te Rau Hingengaro: The New Zealand Mental Health Survey. Wellington: Ministry of Health.

Sadek, N., & Bona, J. (2000). Subsyndromal symptomatic depression: a new concept. *Depression and Anxiety, 12*, 30-39.

Spinhoven, P., Ormel, J., Sloekers, P. P., Kempen, G. I., Speckens, A. E., & Van Hemert, A. M. (1997). A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects. *Psychological Medicine, 27*(2), 363-170.

Sunderland, M., Wong, N., Hilvert-Bruce, Z., & Andrews, G. (2012). Investigating trajectories of change in psychological distress amongst patients with depression and generalised anxiety disorder treated with internet cognitive behavioural therapy. *Behaviour research and therapy, Jun 30;50*(6), 374-380.

The MaGPIe Research Group. (2003). The nature and prevalence of psychological problems in New Zealand primary health care: a report on Mental Health and General Practice Investigation *NZMJ* (Vol. 116, pp. 1-15).

The MaGPIe Research Group. (2006). The treatment of common mental health problems in general practice. *Family Practice, 23*, 53-59.

Thornicroft, G., Chatterji, S., Evans-Lacko, S., Gruber, M., Sampson, N., Aguilar-Gaxiola, S., . . . Kessler, R. C. (2017). Undertreatment of people with major depressive disorder in 21 countries. [Empirical Study; Interview; Quantitative Study]. *The British Journal of Psychiatry, 210*(2), 119-124. doi: http://dx.doi.org/10.1192/bjp.bp.116.188078

Wagner, H., Burns, B., Broadhead, W., Yarnall, K., Sigmon, A., & Gaynes, B. (2000). Minor depresssion in family practice: functional morbidity, co-morbidity, service utilisation and outcomes. *Psychological Medicine, 30*, 1377-1390.

Wang, P. S., Simon, G., & Kessler, R. C. (2003). The economic burden of depression and the cost-effectiveness of treatment. *International Journal of Methods in Psychiatric Research, 12*(1), 22-33. doi: http://dx.doi.org/10.1002/mpr.139

Weisfeld, V., English, R. A., & Claiborne, A. B. (2012). *Public Engagement and Clinical Trials: New Models and Disruptive Technologies: Workshop Summary.*: National Academies Press.

Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., . . . Burstein, R. (2013). Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *The Lancet, Nov 15;382*(9904), 1575-1586.

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica, 67*(6), 361-370.

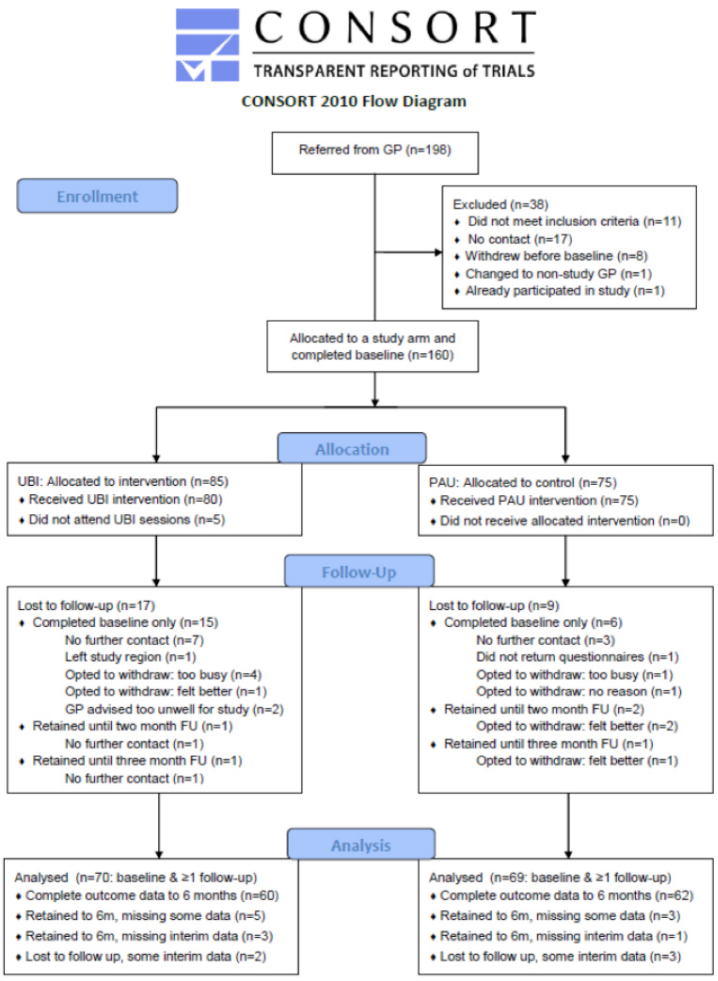


Figure 1. Study flowchart of patient participation.

90x90mm (300 x 300 DPI)

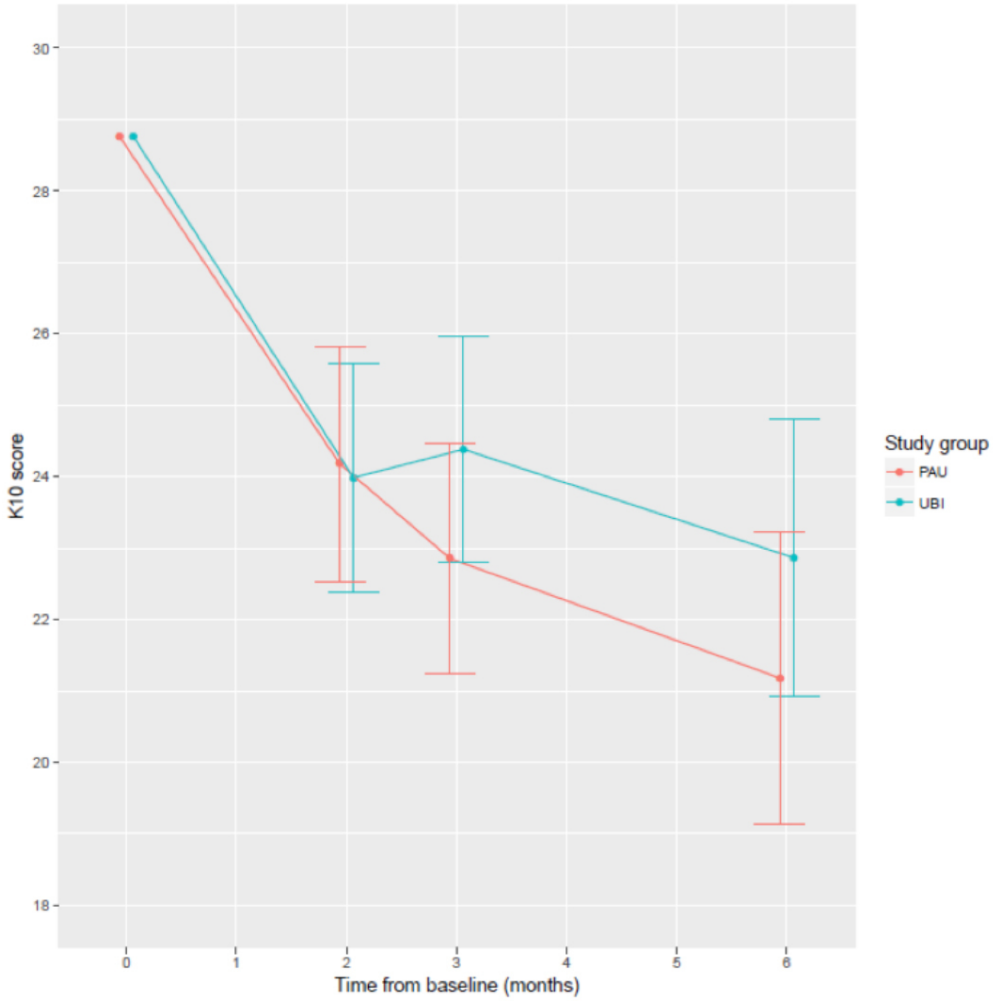


Figure 2. Mean K10 score (95% CI) at baseline and follow up for UBI and PAU study arms.

90x90mm (300 x 300 DPI)

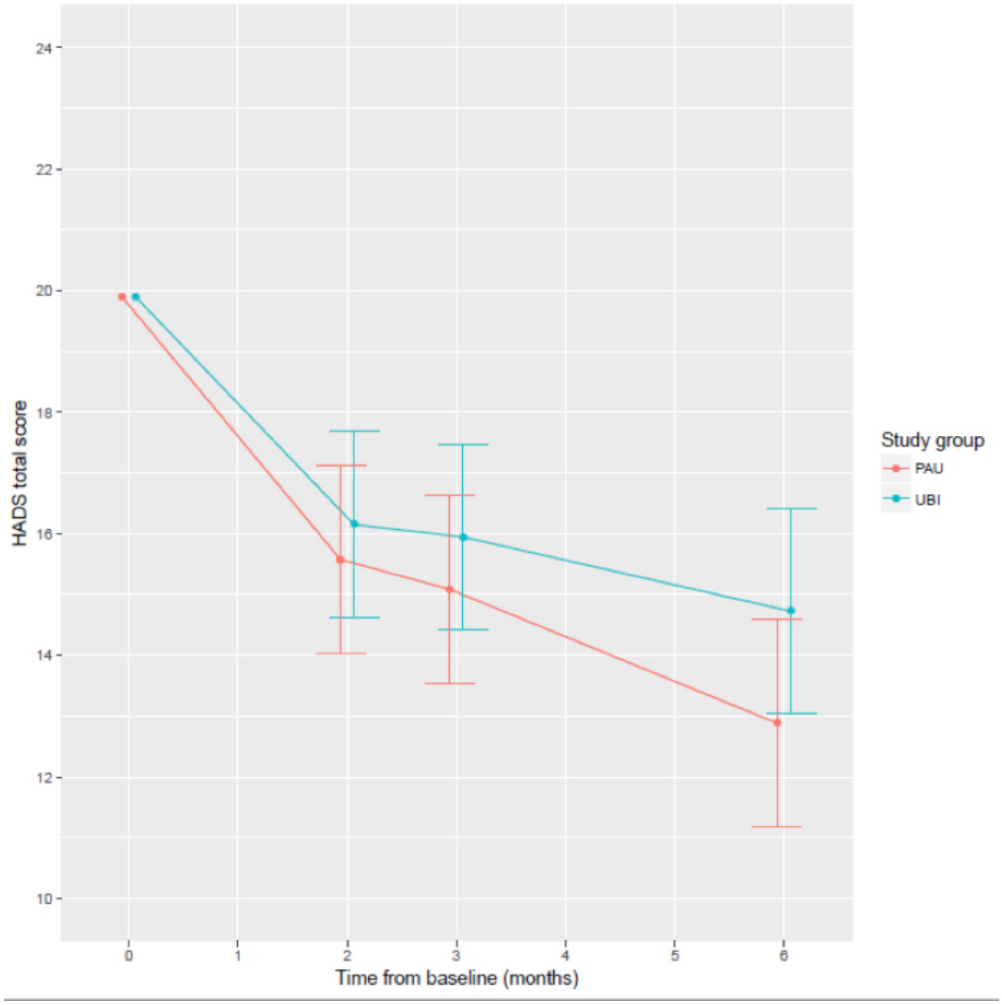
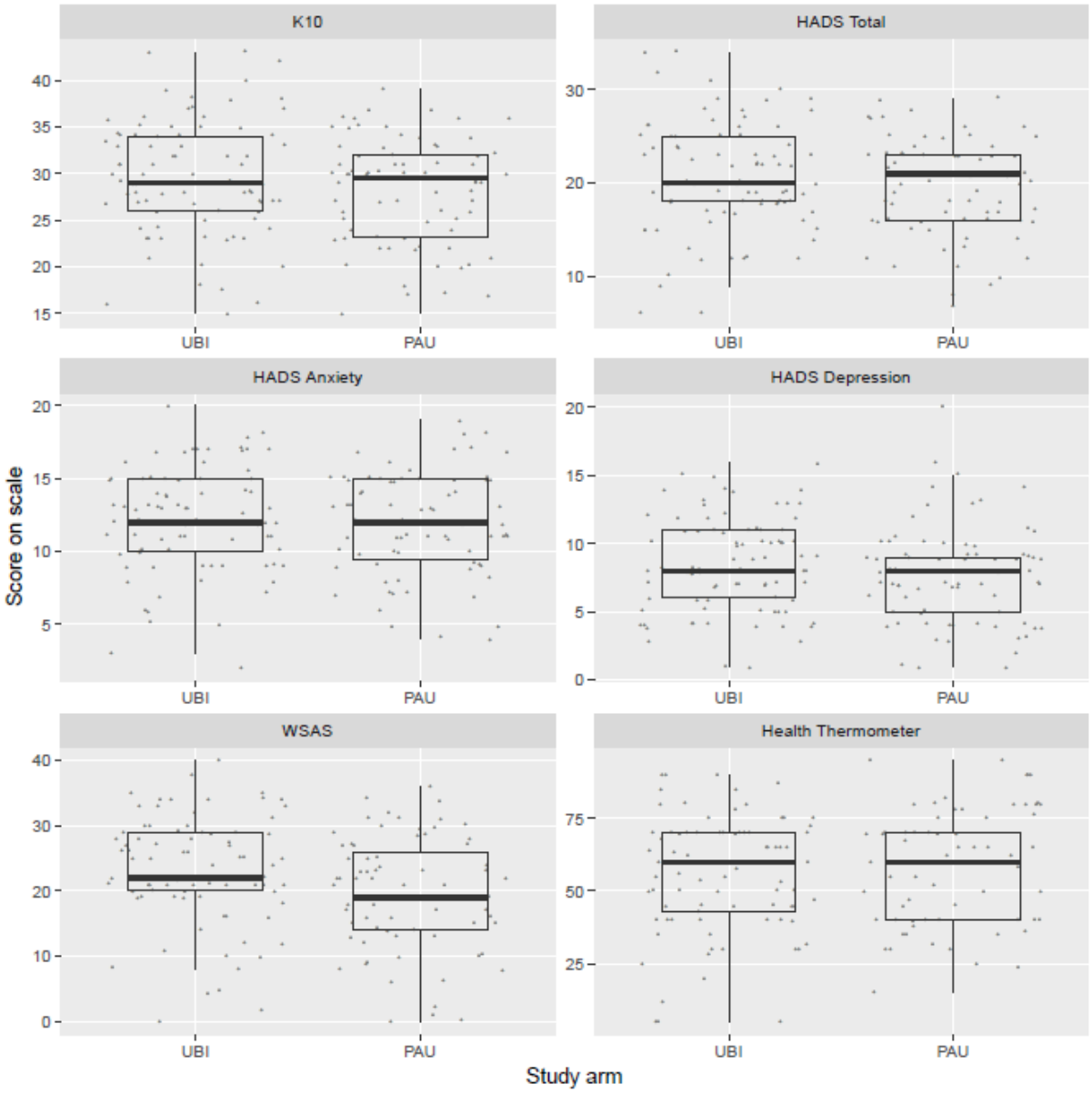


Figure 3. Mean total HADS score (95% CI) at baseline and follow up for UBI and PAU study arms.

90x90mm (300 x 300 DPI)

Supplementary Figure R1. Boxplots of baseline scores for each outcome measure (dots show each individual's score on that measure).



Supplementary Table R1. Number of patients recruited into study by GPs in UBI and PAU study arms.

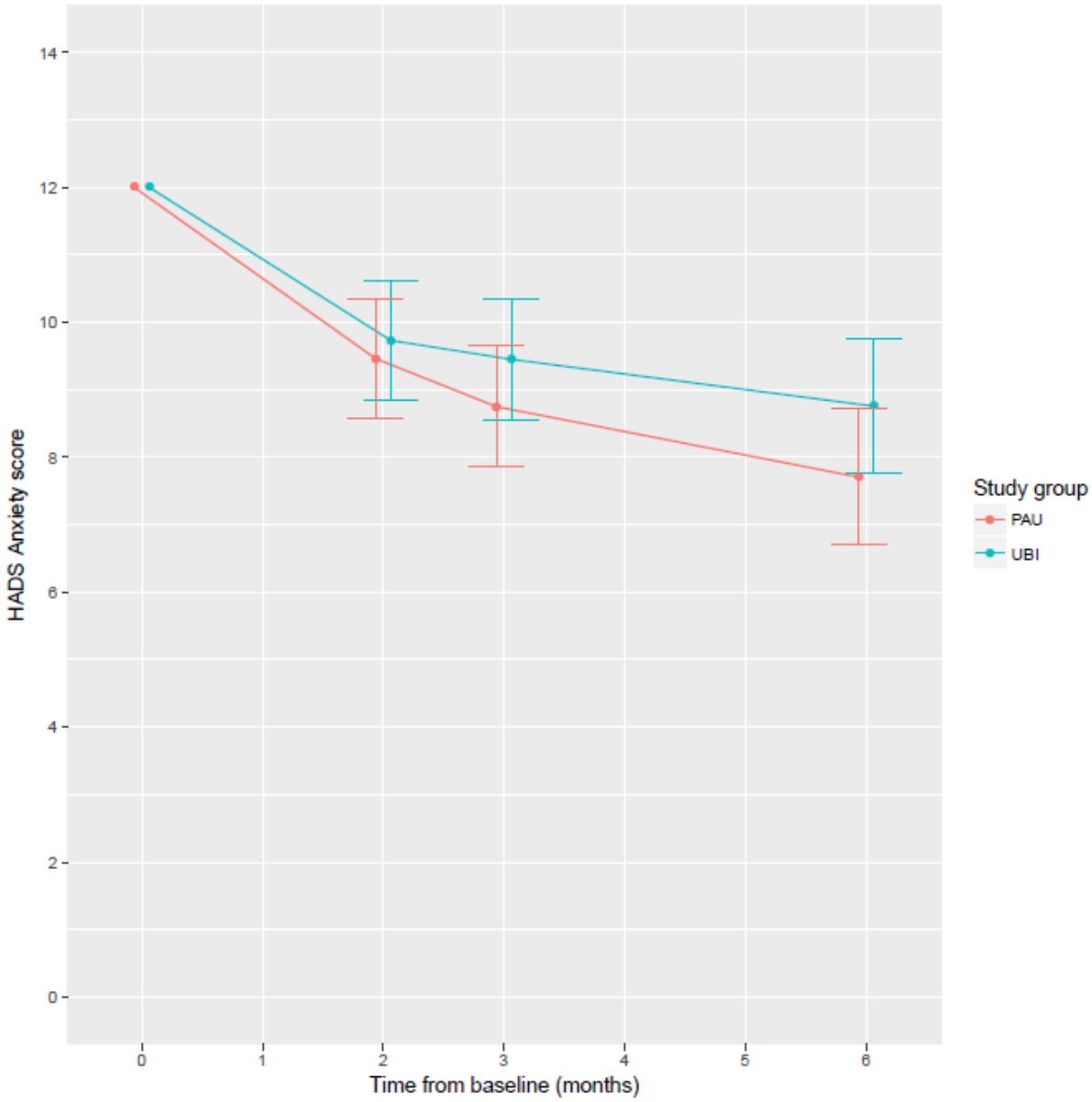
Number of patients recruited by GP	UBI (n GPs*)	PAU (n GPs*)
1	12	8
2	4	2
3	7	5
4	3	0
5	1	2
6	2	0
7	1	0
8	1	1
9	0	2
12	0	1
Total number of GPs	31	21

* Indicates the number of GPs recruiting the stated number of patients (e.g. 12 GPs in the UBI arm recruited one patient each; and five GPs in the PAU arm recruited three patients each).

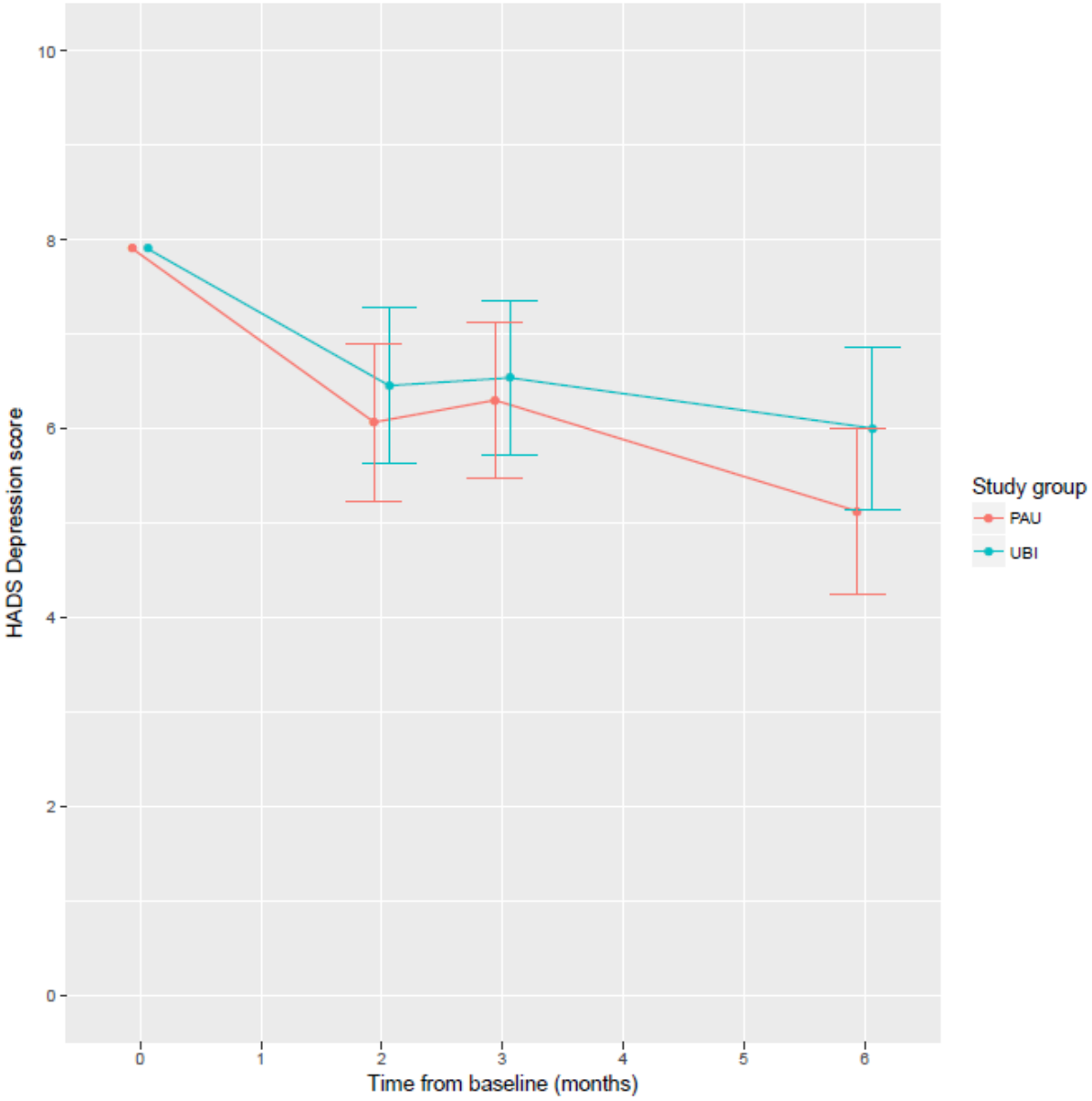
Supplementary Table R2. Mean improvements from baseline to 6 month follow-up for each outcome measure.

Outcome measure	Mean at baseline (both arms)	Mean improvement (95% CI) from baseline to 6 months	
		PAU	UBI
K10	28.8	7.6 (5.5, 9.6)	5.9 (4.0, 7.8)
HADS	19.9	7.0 (5.3, 8.7)	5.2 (3.5, 6.9)
HADS-A	12	4.3 (3.3, 5.3)	3.2 (2.2, 4.2)
HADS-D	7.9	2.8 (1.9, 3.7)	1.9 (1.0, 2.8)
WSAS	21.3	7.7 (5.7, 9.7)	7.2 (5.3, 9.2)
Health Thermometer	57.5	14.0 (9.3, 18.6)	9.0 (4.4, 13.7)

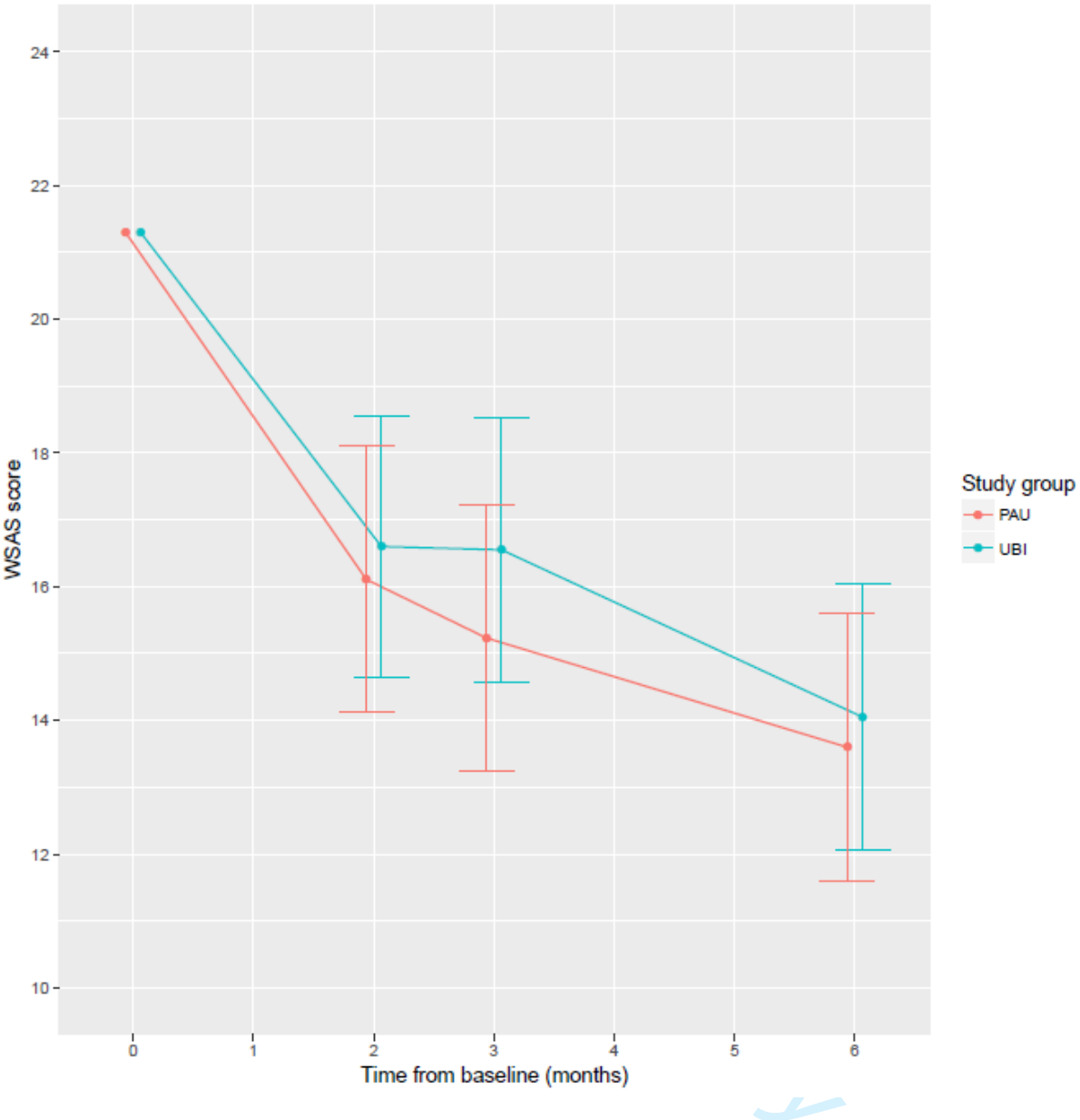
Supplementary Figure R2. Mean HADS Anxiety score (95% CI) at baseline and follow up for UBI and PAU study arms.



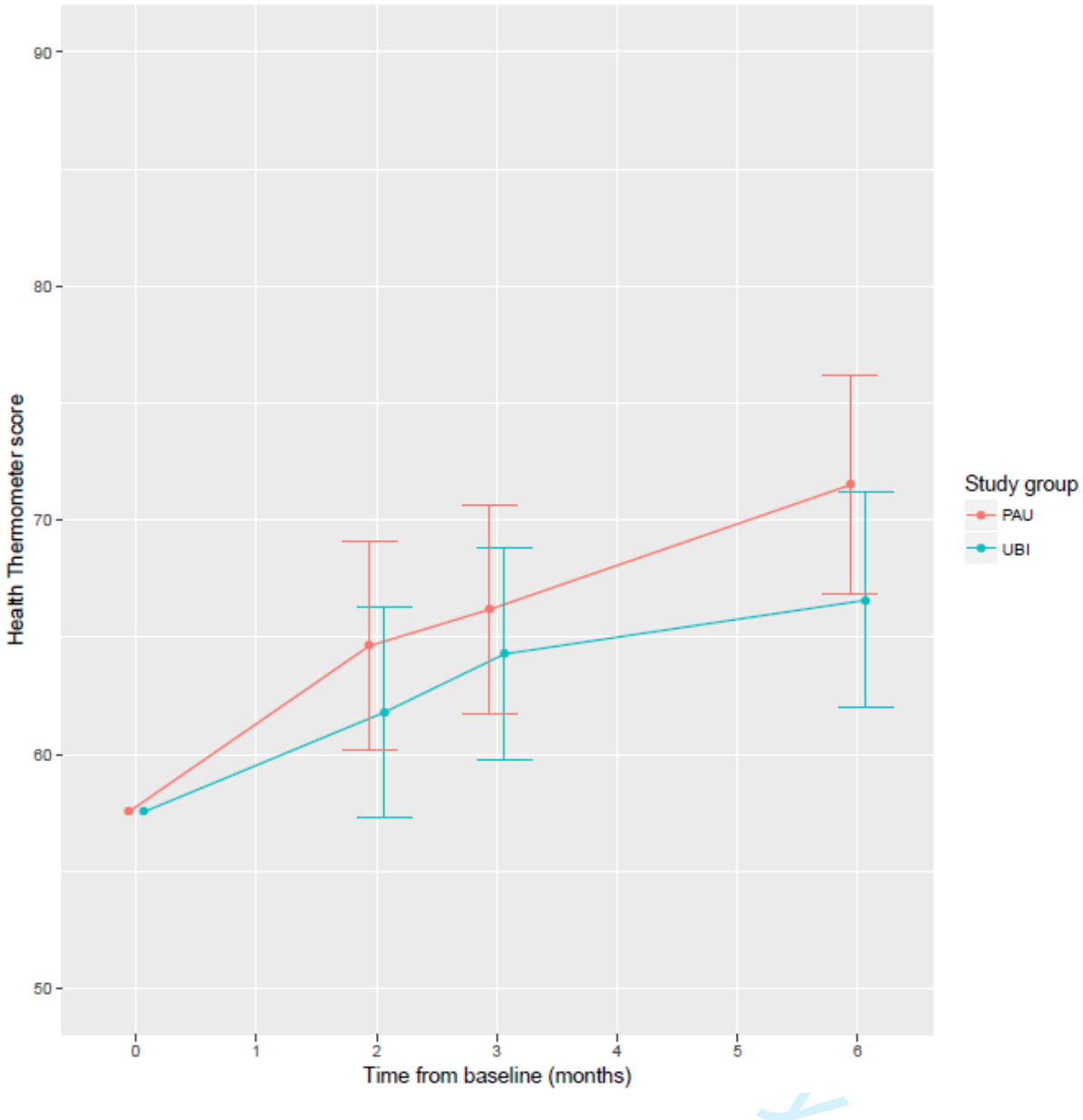
Supplementary Figure R3. Mean HADS Depression score (95% CI) at baseline and follow up for UBI and PAU study arms.



Supplementary Figure R4. Mean WSAS score (95% CI) at baseline and follow up for UBI and PAU study arms.



Supplementary Figure R5. Mean Health Thermometer score (95% CI) at baseline and follow up for UBI and PAU study arms.



Supplementary Table R3. Additional treatment received during UBI trial (from question on 6 month interview)

Type of additional treatment		UBI n (%)	PAU n (%)
Medication status during trial			
	no relevant medication	33 (51%)	34 (52%)
	on medication prior to entering trial	20 (31%)	16 (25%)
	started medication during trial	12 (18%)	16 (25%)
	did not complete question*	20	9
Extended GP consultations (n)			
	0	68 (100%)	46 (71%)
	1-2	0	8 (12%)
	3-5	0	9 (14%)
	6-10	0	2 (3%)
	did not complete question*	17	10
Counselling sessions (n)			
	0	44 (75%)	21 (36%)
	1-2	4 (7%)	13 (22%)
	3-5	2 (3%)	11 (19%)
	6-10	7 (12%)	12 (20%)
	11+	2 (3%)	2 (3%)
	did not complete question*	26	16

* Did not complete 6 month questionnaire and hence no data (UBI n=16; PAUn=9)
Did not answer Meds question at 6 months (UBI: n=4; PAU: n=1)
Did not answer Extended GP question at 6 months (UBI: n=1; PAU: n=1)
Did not answer Counselling question at 6 months (UBI: n=10; PAU: n=7)

Supplementary Methods: Calculation of intra-class correlation coefficients (ICCs) for outcome measures.

ICCs were calculated for each outcome measure in the study to summarise the impact of clustering of outcomes by GPs. These were calculated using simplified mixed linear models with random intercept terms for GPs and no adjustment for covariates. ICCs were calculated in R 3.2.3, using the lme4 package, with their 95% confidence intervals based on 1000 bootstrap resamples calculated using the bootMer() function.

Supplementary Table R4. Intra-class correlation coefficients (ICCs) for each outcome measure in the study.

Outcome measure	ICC	(95% CI)
K10	0.129	(0.045, 0.231)
HADS (total)	0.091	(0.019, 0.189)
HADS Anxiety	0.095	(0.019, 0.198)
HADS Depression	0.142	(0.047, 0.250)
WSAS	0.185	(0.081, 0.308)
Health Thermometer	0.086	(0.013, 0.177)

Reference for lmer package:

Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Table 1: CONSORT 2010 checklist of information to include when reporting a cluster randomised trial

Section/Topic	Item No	Standard Checklist item	Extension for cluster designs	Page No *
Title and abstract				
	1a	Identification as a randomised trial in the title	Identification as a cluster randomised trial in the title	Title page
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)[1,2]	See table 2	In abstract
Introduction				
Background and objectives	2a	Scientific background and explanation of rationale	Rationale for using a cluster design	p. 4-5 also p. 6 (methods)
	2b	Specific objectives or hypotheses	Whether objectives pertain to the the cluster level, the individual participant level or both	p. 4
Methods				
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	Definition of cluster and description of how the design features apply to the clusters	p. 4-5
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons		n/a
Participants	4a	Eligibility criteria for participants	Eligibility criteria for clusters	p.4 (for both clusters and participants)
	4b	Settings and locations where the data were collected		p. 4
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	Whether interventions pertain to the cluster level, the individual participant level or both	p. 7
Outcomes	6a	Completely defined pre-	Whether outcome measures	p. 8

		specified primary and secondary outcome measures, including how and when they were assessed	pertain to the cluster level, the individual participant level or both	
	6b	Any changes to trial outcomes after the trial commenced, with reasons		n/a
Sample size	7a	How sample size was determined	Method of calculation, number of clusters(s) (and whether equal or unequal cluster sizes are assumed), cluster size, a coefficient of intracluster correlation (ICC or *k*), and an indication of its uncertainty	p. 8-9
	7b	When applicable, explanation of any interim analyses and stopping guidelines		n/a (no interim analysis was applied)
Randomisation:				
Sequence generation	8a	Method used to generate the random allocation sequence		p. 6
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	Details of stratification or matching if used	p. 6
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	Specification that allocation was based on clusters rather than individuals and whether allocation concealment (if any) was at the cluster level, the individual participant level or both	n/a
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	Replace by 10a, 10b and 10c	
	10a		Who generated the random allocation sequence, who enrolled clusters, and who assigned clusters to interventions	p. 6 (Recruitment and Randomisation sections)

	10b		Mechanism by which individual participants were included in clusters for the purposes of the trial (such as complete enumeration, random sampling)	p. 7 (Recruitment procedures sub-section)
	10c		From whom consent was sought (representatives of the cluster, or individual cluster members, or both), and whether consent was sought before or after randomisation	p. 6 (for GPs as the clusters) and p.7 (consent for the patients)
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how		GPs unable to be blinded (p 6) Statistician blinded during analysis (p. 9) Research assistant unable to be blinded
	11b	If relevant, description of the similarity of interventions		n/a
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	How clustering was taken into account	p.9-10 Analysis and clustering noted on p. 9
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses		p. 10
Results				
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	For each group, the numbers of clusters that were randomly assigned, received intended treatment, and were analysed for the primary outcome	Clusters (GPs) noted on p. 11, additional detail in Supplementary Table R1. Individual patients noted on p. 11, flowchart in

				Figure 1 (including who was covered in analysis)
	13b	For each group, losses and exclusions after randomisation, together with reasons	For each group, losses and exclusions for both clusters and individual cluster members	Clusters (GPs) covered on p.11 (no losses or exclusions, other than zero-recruitment which is covered in Supplementary Table R1) Patients covered in Figure 1
Recruitment	14a	Dates defining the periods of recruitment and follow-up		p. 5
	14b	Why the trial ended or was stopped		p. 5
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	Baseline characteristics for the individual and cluster levels as applicable for each group	Individual level characteristics reported in Table 1.
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	For each group, number of clusters included in each analysis	Analysis by original assigned groups (methods, p. 9) Number of participants for each analysis: Table 2, Table 3 Number of clusters (across all analyses): Supplementary Table R1
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95%	Results at the individual or cluster level as applicable and a coefficient of intracluster correlation (ICC or k) for each	Effect size and precision given in all tables and figures, and for outcomes

		confidence interval)	primary outcome	reported in body of text ICC reported on p 17 for primary outcomes, and Supplementary Table R4 for all outcomes.
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended		n/a (no binary outcomes used in study)
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory		ICCs reported on page 17 (as noted above) Information on additional treatment received presented p 17
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms[3])		n/a
Discussion				
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses		P 18-19 (recruitment not completed to planned sample size) p 20-21 (other limitations)
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	Generalisability to clusters and/or individual participants (as relevant)	p. 20-21
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence		(across discussion)
Other information				

Registration	23	Registration number and name of trial registry	p.4
Protocol	24	Where the full trial protocol can be accessed, if available	p. 4, reference list for detail
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	p. 22-23

** Note: page numbers optional depending on journal requirements*

Table 2: Extension of CONSORT for abstracts[1,2] to reports of cluster randomised trials

Item	Standard Checklist item	Extension for cluster trials
Title	Identification of study as randomised	**Identification of study as cluster randomised**
Trial design	Description of the trial design (e.g. parallel, cluster, non-inferiority)	
Methods		
Participants	Eligibility criteria for participants and the settings where the data were collected	**Eligibility criteria for clusters**
Interventions	Interventions intended for each group	
Objective	Specific objective or hypothesis	**Whether objective or hypothesis pertains to the cluster level, the individual participant level or both**
Outcome	Clearly defined primary outcome for this report	**Whether the primary outcome pertains to the cluster level, the individual participant level or both**
Randomization	How participants were allocated to interventions	**How clusters were allocated to interventions**
Blinding (masking)	Whether or not participants, care givers, and those assessing the outcomes were blinded to group assignment	
Results		
Numbers randomized	Number of participants randomized to each group	**Number of clusters randomized to each group**
Recruitment	Trial status[1]	
Numbers analysed	Number of participants analysed in each group	**Number of clusters analysed in each group**
Outcome	For the primary outcome, a result for each group and the estimated effect size and its precision	**Results at the cluster or individual participant level as applicable for each primary outcome**
Harms	Important adverse events or side effects	
Conclusions	General interpretation of the results	
Trial registration	Registration number and name of trial register	
Funding	Source of funding	

[1] Relevant to Conference Abstracts

REFERENCES

1 Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, et al. CONSORT for reporting randomised trials in journal and conference abstracts. *Lancet* 2008, 371:281-283

2 Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG at al (2008) CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Med* 5(1): e20

3 Ioannidis JP, Evans SJ, Gotzsche PC, O'Neill RT, Altman DG, Schulz K, Moher D. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med* 2004; 141(10):781-788.

BMJ Open

A cluster randomised controlled trial of a guided self-help mental health intervention in primary care

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-023481.R1
Article Type:	Research
Date Submitted by the Author:	25-Oct-2018
Complete List of Authors:	Mathieson, Fiona; University of Otago – Wellington, Department of Psychological Medicine Stanley, James; University of Otago, Wellington, Public Health; University of Otago, Wellington Collings, Catherine (Sunny) ; University of Otago Wellington, Deans Department Tester, Rachel; University of Otago, Wellington, Primary Health Care and General Practice Dowell, Anthony; Wellington School of Medicine and Health Sciences, General Practice
Primary Subject Heading:	Mental health
Secondary Subject Heading:	General practice / Family practice, Mental health
Keywords:	MENTAL HEALTH, PRIMARY CARE, BRIEF INTERVENTIONS

A cluster randomised controlled trial of a guided self-help mental health intervention in primary care

Fiona Mathieson[1]
Email: fiona.mathieson@otago.ac.nz
James Stanley [2]
Email: james.stanley@otago.ac.nz
Sunny Collings[3]
Email: deanseauow@otago.ac.nz
Rachel Tester [4]
Email: rachel.tester@otago.ac.nz
Anthony Dowell[4]
Email: tony.dowell@otago.ac.nz

* Corresponding author: fiona.mathieson@otago.ac.nz Ph. 64 4 9186034, Fax 64 4 3855877
[1]Department of Psychological Medicine, University of Otago, PO Box 7343, Wellington South 6242, New Zealand
[2]Biostatistical Group, Dean's Department, University of Otago, PO Box 7343, Wellington South 6242, New Zealand
[3]Office of the Dean and Head of Campus Te Tari Manutaki,
Department of Psychological Medicine, University of Otago, PO Box 7343, Wellington South 6242, New Zealand
[4]Department of Primary Health Care and General Practice, University of Otago, PO Box 7343, Wellington South 6242, New Zealand

Word Count: 5801

Abstract

Objectives: To ascertain whether an ultra-brief intervention improves mental health outcomes for patients in general practice with mild-to-moderate mental health concerns.

Trial design: Two-arm cluster randomised controlled trial.

Methods:

Participants: General practitioners (GPs) were invited based on working in a participating general practice. Patients were eligible to participate if aged 18-65, scored ≤35 on the Kessler-10 (K10) and if meeting local mental health access criteria (based on age, low income, or ethnic group).

Interventions: Intervention arm GPs were trained on the ultra-brief intervention (UBI) approach, with participating patients receiving three structured appointments over five weeks. GPs randomised to Practice as Usual (PAU) did not receive training, and delivered support following their existing practice approaches.

Outcome Measures: Primary outcome was patient-level K10 score at 6 months post-recruitment.

Randomisation: GP practices were randomised to UBI training or PAU at the start of the study.

Blinding: GPs were not blinded to group assignment.

Results:

Numbers randomised: 62 GPs (recruiting 85 patients) were randomised to UBI, and 50 to PAU (recruiting 75 patients).

Numbers analysed: 31 GPs recruited at least one patient in the UBI arm (70 patients analysed), and 21 GPs recruited at least one patient in the PAU arm (69 patients analysed).

Outcome: K10 scores from an intention-to-treat analysis were similar in UBI and PAU arms, with a wide confidence interval (mean adjusted K10 difference = 1.68 points higher in UBI arm, 95% CI -1.18, 4.55; p=0.255). Secondary outcomes were also similar in the two groups.

Conclusions: The UBI intervention did not lead to better outcomes than practice as usual. Results from 'negative trials' such as this contribute to the continuing development of brief psychological therapy options for primary care.

Trial registration: Australia New Zealand Clinical Trials Registry ACTRN12613000041752

Funding: Compass Health, Oakley Mental Health Research Foundation, Wellington Medical Research Foundation, University of Otago Research Fund

Strengths and limitations

- Pragmatic effectiveness trial of a mental health intervention in primary care.
- Intervention included Maori cultural adaptations.
- Recruitment issues limit strength of results.
- Intervention was applied to more severe mental health presentations that it was developed for.
- GP degree of adherence to the intervention tool is unclear.

Keywords

Mental Health, Primary Care

Introduction

Mental health is major aspect of health and poor mental health is highly prevalent in the general community. Consistent with international findings, just under 40% of the New Zealand (NZ) population had met criteria for a diagnosable mental disorder during their life, and roughly a fifth had experienced a mental disorder in the previous year [1].

There is also considerable international concern about the healthcare burden arising from mental health problems and substance abuse [2-4]), with the World Mental Health Survey (of 21 countries) suggesting that only 41% of people with depression received treatment that met even minimal standards [5].

In NZ, as in other OECD countries, mental health problems are common presentations in primary care. Around one-quarter of primary care patients (26.5% and 29.8% of men and women, respectively), attending their general practice in NZ met criteria for a mild-moderate mental health disorder and an estimated 50-70% of mental health concerns are managed exclusively at the primary care level, since secondary care services have become more targeted towards severe and enduring mental illness in recent years [6] .

Internationally there is a call for psychological therapies to be more widely available in primary care [7], and growing unease about increasing levels of antidepressant medications being prescribed compared with the limited resources available for psychological interventions [8]. However, treatment options at the primary care level are limited, with GPs expressing concerns about gaps in services for patients with mild-moderate mental health presentations and a desire to offer a brief intervention themselves [9]. In NZ, GPs reported that as few as 22% of patients with mild-moderate mental health syndromes receive any formal help [10].

Such patient presentations often comprise sub-threshold syndromes [11, 12], and cases of mild-moderate common mental disorder. These are combinations of problems such as anxiety, depression, substance use and interpersonal problems that do not meet the threshold for disorder

in standard diagnostic systems such as DSM-5. Often these arise in the context of social problems or family or economic stress. In NZ, 36% of general practice attendees report anxiety, depression or substance-use, or a combination of these issues [6]. Such presentations can be associated with significant impairment in functioning and suffering [13, 14], with some going on to develop severe depression [15, 16]. Intervention may be warranted for up to 80% of those affected [10, 13], but referral out of the practice can be problematic due to referral eligibility criteria, waiting times, administrative issues and cost [9, 17, 18].

Increasing knowledge of the burden of mild-moderate disorder led to the development of a platform of Primary Mental Health Initiatives in NZ, which included some increase in access to psychological therapies and extended consultations with GPs. The inclusion criteria for these initiatives, however, mean that only up to 15% of the population can gain access to those services [9].

This service-gap led us to develop a GP delivered ultra-brief intervention (UBI), with development and refinement based on service user feedback [19]. This model has the advantages of avoiding the need for referral on to an expensive professional, such as a psychologist, of being easily accessible to patients, and of potentially building on existing trusted relationships. This fits with the movement towards alternative methods of service delivery for mild to moderate mental health presentations, often termed 'low intensity' interventions. These interventions often include guided self-help, bibliotherapy and computerised delivery of care, with current evidence suggesting that even minimal therapist contact leads to better outcomes than self-help alone [20-23].

UBI was feasibility tested with a group of 16 patients and then adapted for Maori (the indigenous people of New Zealand) and feasibility tested with a group of 9 patients [24, 25]. Based on questionnaire feedback, clinician and patient satisfaction ratings for both feasibility studies were very positive in terms of relevance and acceptability. The psychological well-being of the patients, as measured by the Kessler-10 (K10) [26], was also significantly improved post-intervention (at 3 month follow-up) for both Maori and non-Maori, although there was no control group [24, 25]. Based on these initial findings we designed a cluster randomized controlled trial to measure the effectiveness of UBI.

The aims of the study were to compare patient-level outcomes on (1) mental health state (as measured by K10 scores) at 6 months between UBI and practice as usual (PAU) study arms (primary outcome) and (2) levels of distress (depression and anxiety) and functioning (work, social and relationship) at 8 weeks and 3 months between UBI and PAU study arms (as secondary outcomes).

Methods

A protocol for this study has been previously published, and includes description of planned analyses [27]. The trial was registered prior to recruitment commencing with the Australia New Zealand Clinical Trials Registry (registration ACTRN12613000041752.)

Design

We used a pragmatic two-arm single blinded, cluster randomised controlled trial of UBI compared with PAU, in a primary care setting. GPs were randomised by practice to exclusively deliver either UBI or PAU to all their recruited patients. GPs were treated as the clusters in the study design (while there was clustering by practice, the GPs were treated as the unit of analysis as practitioner attributes were anticipated to be a higher source of variability in outcomes.) Analysis followed an intention-to-treat approach.

Setting

The study was conducted in general practices in the greater Wellington region, New Zealand. This included practices in both city and semi-rural settings, serving populations from a wide range of socio-economic backgrounds. Recruitment took place between 1/5/2013 and 1/7/2016. The trial ended prior to achieving the final sample size when funding for data collection was exhausted.

Participants

This was a pragmatic trial supported within existing treatment services. GPs were eligible to participate if they were currently working in a practice that was part of the Compass Health Primary Health Organisation (PHO) which covers the greater Wellington region.

Patients were eligible if aged between 18 and 65 and identified by their GP in a routine appointment as experiencing stress or distress. Patients were required to score 35 or less on the Kessler Psychological Distress Scale (K10) [26, 28] during their initial GP consultation, with no lower cut-off on this score. The present study followed previous study protocols [24, 25] by including scores between 30 and 35 on the K10 as indicative of mild to moderate levels of psychological distress rather than major psychiatric disorder. Individuals taking anti-depressant or other psychiatric medications were eligible to participate in the study.

Patients were excluded if they lacked fluency in English (as the intervention is an English-language based 'talking therapy'); had significant levels of cognitive impairment as determined by the GP; or had reported recent or acute suicidal ideation (i.e., within the previous 2 weeks). Chronic low level suicidality did not exclude an individual from participating. However, GPs were informed of patients who had high scores or suicidality at screening, or for whom referral to appropriate (secondary) mental health services by GPs was indicated, and these patients were not eligible to participate further in the study.

Inclusion criteria were based on the access criteria of a local partner primary health organisation (PHO) to psychological therapies. These criteria were youth (defined as 18-24 years old), or individuals aged 25 years or older with low income, or Māori or Pacific Island heritage.

Recruitment of practices and GPs

Initial recruitment of practices was supported by the partner PHO. GPs were identified using primary health organisation and practice lists. All of the practices contracted under the partner PHO were contacted (N=52) and invited to participate in the study, and an effort was made to contact all of the GPs within these practices by email, telephone or in person. A total of 23 practices initially consented to participate in the study and a further 18 were recruited during the course of the study. Two practices merged and three withdrew (in each case the single participating GP left the practice) leaving a total of 37 practices involved in the study.

Randomisation of practices to study arms

Consenting practices were randomised to provide either UBI or PAU to eligible patients. Randomisation was conducted at the practice level to reduce the risk of contamination if GPs from the same practice were assigned to opposite study arms. To ensure approximately equal numbers of GPs per study arm, randomisation of practices was conducted within five strata, according to the number of participating GPs (one/two/three/four/more than four). An additional two practices dedicated to youth health that were not part of the partner PHO were included and randomised into each arm of the study (i.e. these two practices formed their own stratum). Practices were entered into the trial following consent from individual participating GPs in that practice. Randomisation of all consenting practices was conducted following this step by the project biostatistician (JS) using a computer-based randomisation following the above stratification profile.

GPs randomised to the UBI study arm completed a single two-hour training session (as previously described [25]). Due to the training nature of the intervention, it was not possible to blind GPs as to their study arm allocation.

Recruitment procedures

GPs identified patients with common mental health problems who might fulfil study criteria during routine appointments. These patients were screened by the GP for eligibility (using the K10), and referred to the study team. A research assistant then contacted potential participating patients, met with them in person where possible to explain the study, confirm eligibility, obtain consent to participate, and collect pre-treatment (baseline) data. Measures were then collected by mail or email at post-treatment (8 weeks, 3 months and 6 months). Patients received compensation (NZ $30 [US$21] vouchers, and entry into a draw for an iPad) following the completion of the final questionnaire, to recompense for time and effort in participating in the study.

Intervention

UBI is a low intensity self-management programme which can be delivered by a GP after a single two-hour training session using a treatment manual based on structured problem solving, motivational interviewing and cognitive behaviour therapy (supported with self-help booklets on relationships, bodily stress, breaking habits and stress management).

Patients who consented and completed the intake data collection (K10 and baseline measurements) received the GP-led intervention in three short, structured face-to-face sessions (one 30 and two 15 minute sessions) over a five to six week period. Relevant booklets were provided to the patient after the first session, to be used in the following session. In New Zealand a stepped care approach to management guides the practitioner towards using the most appropriate therapy option for the severity of presentation. UBI was designed for mild to moderate presentations and in training GPs were comfortable with the use of the UBI approach for first line management. The study protocol allowed for patients in either study arm to alter their treatment as needed (e.g. access other talking therapies, or commence mental health medications). Patients were blinded as to their study allocation in that patients in PAU practices were not informed that the UBI was offered in practices randomised to deliver UBI. They were simply told that the study was looking at the effectiveness of PAU [27].

Practice as usual

Patients in the PAU study arm received GP support delivered according to their practice as usual (and available existing services). PAU typically consists of supportive counselling in a 15 minute face-to-face consultation, the provision of psychotropic medication, referral to psychological or other counselling options, or referral to relevant community services.

Patient characteristics

Patients are described on the basis of age, gender, prioritised ethnicity and NZiDep, a NZ-developed index [29] of individual-level socioeconomic deprivation.

GPs in practices assigned to the PAU study arm received optional training in the intervention at the end of the study.

Patient and Public Involvement

This study had input from an academic mental health consumer (i.e. an academic who is also a mental health service user and who conducts research from a service user perspective) as part of the research team at the feasibility stage, and designed the intervention based on feedback from a focus group process with potential patient users of the mental health intervention which asked what characteristics such an intervention would need to have. This collaborative process is fully described in [19]. This RCT did not have academic consumer or patient involvement in the recruitment to and conduct of the study and the burden of the intervention was not assessed by the patients. Results of this study will be disseminated by email to GP participants who indicated they wanted them on the consent form.

Outcome measures

The primary outcome measure was the K10 scale [26, 28] score at 6 months (adjusted for score at baseline: see analysis). The K10 is widely used as a clinical outcome measure in Primary Care and General Practice in NZ [9]. A 6 month follow up period was chosen to obtain a sufficient period of assessment following the end of the intervention while at the same time balancing out challenges in patient cohort retention. All analyses were conducted to look at patient-level outcomes.

Secondary outcomes were:

1) Hospital Anxiety and Depression Scale (HADS), which measures the severity of depressive and anxiety symptoms in outpatient hospital settings [30]. Reductions in HADS score indicate reduced anxiety and depression.
2) Comparison of K10 scores by treatment group at 8 weeks and 12 weeks, adjusted for baseline scores (to capture short and medium term effectiveness).
3) Work and Social Adjustment Scale [31], a measure of work, social and relationship functioning) administered at baseline, 8, 12 and 26 weeks.

Outcomes were measured at the same time points in both UBI and PAU groups (baseline, and at 8, 12, and 26 weeks following baseline)

Statistical methods

Sample size and Power analysis

Sample size for the cluster randomised trial was calculated using a simulation method, using standard deviations of patient outcomes from the UBI feasibility study (standard deviation of post-treatment scores = 7.5 [25]). To detect a difference in K10 improvement scores of 6 points in the UBI arm compared with 2 points in the control arm (at 80% power and alpha = 0.05) would require 15 GPs per arm recruiting eight completing patients each on average (n=240 total with complete data). Adjusting for loss to follow-up of 20% gave a recruitment target of ten patients per GP. The simulation settings roughly correspond to an intraclass correlation (ICC) of 0.15 for considering clustering of patient scores by GP (equivalent to the ICC from the feasibility study [25]). Power analysis for the secondary HADS outcome indicated 80% power to detect a difference of 3.2 points between groups (based on a standard deviation of approximately 6 [32]) assuming a similar ICC for the HADS scale as for the K10 measure (empirical data were not available).

Data Analysis

The statistician was blinded to the intervention or control status of participants (both practices and patients) during conduct of the study and analysis. Results were unblinded once analysis was complete. Data processing and analysis were conducted in R 3.2.3 (R Institute, Vienna) with linear mixed models fit using the lmer package [33] and imputation conducted using the mice package [34].

For the primary outcome, K10 scores at 6 months were compared between the intervention and control groups using mixed linear models (comparing post-intervention scores between groups, adjusting for intake score as a covariate, and treating GP clusters as random slope effects). Analysis was conducted on an intention-to-treat basis according to the study arm for each patient at entry into the study. Analyses were adjusted for all other baseline covariates (age, gender, ethnicity, educational level, and NZiDep). The original protocol stated that analyses would only be adjusted for baseline-values of each score: given some slight imbalance in sociodemographic

characteristics it was decided to adjust for other baseline covariates in the main analyses. The originally planned analyses are presented in supplementary materials (overall patterns discussed in the body of the results).

Missing data were handled through the mixed linear models approach to the data, which allows for patients with missing data on the final outcome to be included in analyses, which in effect estimates a final outcome value conditional on the observed data at other follow-up times (i.e. validity being predicated under the assumption that the missing observations are missing at random [MAR], conditional on the observed data [35, 36]). Participants missing all follow-up data were excluded from this main analysis. The null hypothesis for this test was that the K10 scores at 26 weeks (adjusted for baseline score) were not different for the intervention and control groups.

Sensitivity analysis for missing follow-up data in the K10 primary outcome were planned and conducted following completion of the main analysis, and hence were not noted in the trial registration or protocol paper. These analyses covered two scenarios: firstly, an analysis with multiple imputation of missing outcomes, conditional on observed baseline sociodemographics and baseline outcome data. This analysis hence included participants who only had baseline data recorded (excluded from the main mixed models analysis), and assumes that the unobserved outcome data are missing at random conditional on observed data: that is, that individuals who were missing from all follow-up data collections had the same outcome profile (on average) as participants with similar profiles at baseline [37]. The second sensitivity analysis explored this missing at random assumption: those missing data post-baseline were (i) assumed to have scores at 6 months that were four points worse than their imputed score in the first sensitivity analysis; (ii) assumed to have had no improvement from baseline (last observation carried forward); and (iii) assumed to have had poorer outcomes at six months than at baseline (4 points worse than baseline). Full details of the imputation procedure and sensitivity analyses are presented in the Supplementary material, and results are summarised and discussed in the main body of the results and discussion.

For the secondary analysis, differences in mean scores on the K10 outcome were reported at 8 weeks and 3 months (using the same methods as above, within the mixed linear models

framework). Analysis of the HADS and WSAS scores at 8 weeks, 3 months and 6 months utilised the same methods as for the K10 outcome. Analysis of outcomes at 8 weeks and 3 months was not specified on the clinical trials registry, but was noted in the previously published protocol paper [27].

The EQ-5D-3L was noted as a secondary outcome for quality of life in the trial registry. This measure was intended as part of an economic analysis that was not implemented, and no other economic data was collected as part of this study.

Intra-class correlation coefficient (ICC) values were calculated for each outcome measure as a summary of clustering according to GPs. Because our analytical models only accounted for clustering at the level of individual GPs, we also examined ICC values when clustering was considered as a multilevel structure (GPs nested within specific practices). Details of the calculation methods are provided in the Supplementary Materials.

Additional treatments received during the trial (including medication and talking therapies) were analysed by study arm, based on self-report data collected at the 6 month follow-up. This descriptive analysis was not specified in the study protocol.

Confidentiality and data management

Consenting patients had their rights explained along with provision for data confidentiality. Paper and digital copies of the data were secured in locked storage on the premises of the University of Otago, Wellington. The questionnaire data was de-identified and entered into a spreadsheet for subsequent analysis.

Ethics approval

Ethical approval was received from the Health and Disability Ethics Committees (HDEC), Ministry of Health (Northern B Health and Disability ethics committee 12/NTB/2).

Adverse events were not anticipated in this trial, and arrangements were made to feedback clinical information to GPs if deemed necessary (e.g., high K10 scores or concerning self-reported statements about a patient's safety) in the course of data collection.

Results

GP Participants

A total of 41 practices agreed to participate, with a total of 112 individual GPs consenting to take part in the study (n=62 for UBI, and n=50 for PAU). Of these GPs, 31 recruited at least one patient into the study in the UBI arm (from 22 practices), and 21 recruited at least one patient in the PAU arm (from 12 practices). The numbers of GPs recruiting different numbers of patients is shown in Supplementary Table R1.

Patient Participants

Figure 1 summarises the flow of patients into the study and participation in the interventions and follow-up. A total of 198 patients were referred into the study, and 160 met eligibility criteria and completed baseline assessments. The vast majority of these completed at least one post-intervention follow-up (70 / 85 in the UBI arm [82%]; and 69/75 in the PAU group [92%]) and hence contributed to the data analysis. These patients represented 29 GPs (from 21 practices) and 20 GPs (from 12 practices) in the intervention and control arms respectively.

<Insert figure 1 about here>

Baseline data

Baseline sociodemographic characteristics of patients are presented in Table 1 for the two study arms. The two groups were roughly comparable at baseline, with a few more male participants and a slightly younger age profile in the UBI arm, but with a greater representation of females in the study overall.

Table 1. Patient sociodemographic profile by study arm.

Variable	Level	Study Group	
		UBI (Total n = 85)	PAU (Total n = 75)
		n (%)	n (%)
Gender			
	Female*	56 (65.9)	57 (76.0)
	Male	29 (34.1)	18 (24.0)
Age Group			
	15-24	55 (64.7)	37 (49.3)
	25-34	16 (18.8)	15 (20.0)
	35-44	3 (3.5)	13 (17.3)
	45-54	5 (5.9)	6 (8.0)
	55+	6 (7.1)	4 (5.3)
Ethnicity			
	NZE/Other	61 (71.8)	54 (72.0)
	Māori	19 (22.4)	14 (18.7)
	Pacific	4 (4.7)	2 (2.7)
	Asian	1 (1.2)	5 (6.7)
Highest education			
	At least secondary	78 (91.8)	71 (94.7)
	No secondary level	7 (8.2)	4 (5.3)
NZiDep			
	0 (least deprived)	18 (21.2)	11 (14.7)
	1	16 (18.8)	17 (22.7)
	2	15 (17.6)	11 (14.7)
	3	10 (11.8)	10 (13.3)
	4	9 (10.6)	12 (16.0)
	5 (most deprived)	17 (20.0)	14 (18.7)

* Includes one individual self-identifying as Female (transgender)

Mean baseline scores on the outcome measures were also similar between the two groups (Table 2, showing means and standard deviations). Boxplots of the distribution of baseline scores on each outcome scale are given in Supplementary Figure R1.

Table 2. Mean (standard deviation) of baseline scores for outcome measures by study arm

Outcome variable	Study Group	
	UBI (Total n = 85)	PAU (Total n = 75)
	mean (sd)	mean (sd)
K10*	29.5 (6.2)	28.1 (5.7)
HADS – total	20.6 (5.9)	19.5 (5.1)
HADS – anxiety	12.1 (3.6)	11.9 (3.5)
HADS – depression	8.5 (3.5)	7.7 (3.6)
WSAS	23.0 (8.2)	19.6 (8.5)
Health Thermometer**	55.4 (19.9)	58.8 (18.7)

* One patient in PAU group missing baseline value.
** Higher scores on the health thermometer indicate better health.

Health Outcomes at Follow-up

For the K10 primary outcome at 6 months the mean difference for UBI compared to PAU arm favoured the PAU arm (mean adjusted difference = 1.68, 95% CI -1.18 to 4.55; p = 0.255: adjusted for age, gender, ethnicity, educational level, and NZiDep), as shown in Table 3 (where positive differences indicate a better outcome for the PAU than UBI arm) While this result indicated no significant difference in K10 scores between the UBI and PAU arms (see Figure 2), each group had a reasonable improvement in K10 score from baseline (see Supplementary Table R2: for the PAU group mean improvement = 7.6, 95% CI 5.5, 9.6; and for the UBI group mean improvement = 5.9, 95% CI 4.0, 7.8).

Table 3. Mean difference in primary and secondary outcomes (difference in change relative to baseline)

	Mean adjusted difference (UBI minus PAU)*					
	8 weeks		3 months		6 months	
Outcome variable	mean diff (95% CI)	p	mean diff (95% CI)	p	mean diff (95% CI)	p
Primary outcomes**						
K10	-0.19 (-2.55, 2.16)	0.872	1.53 (-0.79, 3.84)	0.203	1.68 (-1.18, 4.55)	0.255
HADS	0.57 (-1.68, 2.82)	0.620	0.86 (-1.38, 3.10)	0.456	1.85 (-0.62, 4.31)	0.149
Secondary outcomes**						
HADS-A	0.27 (-1.02, 1.56)	0.684	0.70 (-0.60, 2.00)	0.296	1.05 (-0.39, 2.50)	0.161
HADS-D	0.39 (-0.82, 1.60)	0.533	0.24 (-0.96, 1.44)	0.701	0.88 (-0.38, 2.14)	0.178
WSAS	0.49 (-2.40, 3.38)	0.740	1.32 (-1.58, 4.22)	0.377	0.45 (-2.47, 3.37)	0.762
Health Thermometer	2.84 (-3.64, 9.31)	0.395	1.90 (-4.59, 8.39)	0.569	4.93 (-1.77, 11.62)	0.156

* Positive differences indicate better improvement in PAU than UBI arm, adjusted for baseline value of score and age, gender, ethnicity, educational level, and NZiDep.

** Number of participants contributing data to each analysis: UBI n = 70, PAU n = 69 (except for K10: PAU n = 68)

<insert Figure 2 about here>

The mean adjusted difference on the HADS measure at 6 months between UBI and PAU measures was 1.85 (95% CI = -0.62, 4.31, p = 0.149; see Table 3), though both groups again showed an improvement in mean score from baseline (Supplementary Table R1). Mean scores at each follow-up time are presented in Figure 3.

<insert Figure 3 about here>

Similarly, for all secondary outcome measures (HADS Anxiety and Depression sub-scales, WSAS, and Health Thermometer), the adjusted difference in outcomes at 6 months showed no significant advantage for either UBI or PAU measures (with relatively broad confidence intervals for these differences: see Table 3.)

Estimates of secondary analyses of outcomes at earlier follow-up times (8 weeks and 3 months) are also presented in Table 3. Differences between UBI and PAU were generally most pronounced at the final follow-up (6 months) compared to the interim follow-ups. Trajectories for mean scores in each group are presented in Supplementary Figure R2, Supplementary Figure R3, Supplementary Figure R4 and Supplementary Figure R5.

Ancillary analyses

Supplementary Table R3 presents information on types of additional treatment received for those who completed the 6-month follow-up assessment (summary not specified in protocol). Similar proportions of completing patients between study arms were either on medication for mental health condition(s) at the beginning on the trial (UBI = 31%; PAU 25%), or started medication during the trial (UBI=18%; PAU=25%). Access to extended GP consultations or counselling sessions was higher for the PAU arm than for UBI (no UBI patient had an extended GP consultation, compared to 29% of PAU patients; and 25% of UBI patients had one or more counselling sessions, compared to 64% of PAU patients.)

Intra-class correlation coefficients (ICCs) for the outcome measures are presented in Supplementary Table R4. For the K10 (ICC = 0.129, 95% CI 0.045 – 0.231) this was relatively close to the ICC values used in planning the sample size for the study. We also examined clustering effects for GPs as nested within GP practice clusters: this additional complexity (not implemented in our main analytical models) had little impact on ICCs for the K10 or HADS measures, though it did suggest slightly higher ICCs (greater clustering of outcomes than considering GPs alone) for the WSAS and Health Thermometer.

We also conducted three sensitivity analyses for our primary outcome of K10 scores at 6 months. These analyses are described in more detail in the Supplementary Methods and Results.

The first sensitivity analysis used the same linear mixed models analysis as the main reported analysis, but adjusted only for baseline values of the outcome score (as specified in the original protocol: no adjustment for other baseline covariates). This returned a slightly smaller mean difference between study arms (again with a poorer mean K10 score in UBI compared to PAU: difference = 1.07, 95% CI -1.67, 3.82; p=0.447) but does not control for the covariate imbalance seen in recruited participants (as shown in Table 1).

The second and third sensitivity analyses both aimed to consider the impact of loss-to-follow-up on the primary outcome analysis, assuming data were missing at random (MAR) or missing not at random (MNAR). Full details of implementation are in the Supplementary Methods. Both analyses include all randomised participants. An initial table gives the baseline covariates for those with and without follow-up in the PAU and UBI groups (Supplementary Table R5).

The analysis of outcomes under an MAR assumption (including all randomised participants) was almost identical to the main results (Supplementary Table R6). Analyses of outcomes under MNAR assumptions were also not substantively different from the main results (Supplementary Table R7): the most conservative result returned a mean difference of 2.03 points on the K10 (95% CI -0.63, 4.70: Scenario 1 in Supplementary Table R7) which was slightly bigger than the mean difference seen in the main results (1.68 points, as per Table 3).

Discussion

The brief psychological treatment (UBI) delivered by GPs in New Zealand in routine practice settings did not lead to better outcomes than practice as usual (PAU) in this pragmatic efficacy trial, with the point estimate for the primary outcome favouring PAU over UBI.

UBI appeared to be slightly less effective than PAU in reducing distress as measured by the K10 (though the difference was not statistically significant). The K10 was originally introduced as an assessment measure of psychological distress, but has also been used to track change in mental health status following intervention [38]. There were no significant differences in the secondary measures either.

We were unable to achieve full recruitment to match the pre-determined sample size: the study recruited 160 eligible participants across both study arms, against our target of 240 participants with complete data. As such, we were unable to rule out non-inferiority of the intervention (UBI) compared to PAU in reducing the disability and distress associated with mild to moderate mental health problems: the bounds of the confidence intervals for the two main outcomes (K10 and HADS measures) included sizable-magnitude better outcomes for PAU over UBI (e.g. the upper bound for the K10 was a 4.55 point advantage for PAU).

Both UBI and PAU arms showed improvement in clinical outcome over the 6 month course of the study. These findings are in keeping with other work which demonstrates clinical effectiveness of brief psychological interventions in primary care settings [39].

These results suggest that GPs in both arms were achieving clinical benefit. We cannot rule out that UBI performs slightly worse than PAU, but our results are inconclusive due to our reduced sample size. For the last 10-20 years in many OECD jurisdictions there has been a focus on improving mental health care provision in primary care settings. In New Zealand this has taken the form of the introduction of locally based primary mental health initiatives, which have increased access to psychological services and provided opportunity for increased engagement (and remuneration) by General Practitioners to undertake mental health consultation work [9]. These opportunities were available to the PAU, and may partially explain the relative success of this ‘control’ arm in the study.

Strengths of this study

We consider the results of this trial a useful addition to the literature for two reasons. Firstly they describe the introduction of potentially useful adjuncts to existing therapy approaches in primary care in a randomised controlled setting, and secondly the 'negative results' raise questions about the challenges of conducting pragmatic trials of psychological interventions in primary care and also about the nature and effectiveness of PAU treatments. Feedback received from GPs during the training sessions suggested that elements of the UBI such as active listening, goal-setting; making a specific plan and following up on it are already used in routine practice. UBI had previously been piloted and shown to be both feasible and acceptable to both clinicians and patients in a general practice setting [25]. It was also able to be adapted in a culturally responsive way [24]. During the course of the trial and following its completion there has been significant interest expressed by both patients and GPs in obtaining copies of the booklets and using elements of the UBI approach in routine consultations. Verbal feedback suggests that GPs particularly liked the helpful/unhelpful behaviour chart which was used to discuss how problems were maintained, the explicit linking of emotional responses to physical symptoms and the use of commitment and capability rulers (a motivational interviewing strategy).

There is an active debate about the optimal balance of intervention components for the management of common mental health problems, with an increasingly varied range of options available. Patients potentially have access to traditional face to face intervention with a therapist, access to materials available on the internet, and further access to rapidly developing telemedicine and virtual consultation options [40, 41]. Our study shows that over the course of the trial, patients and GPs were able to adapt the standard pattern of the GP consultation to a series of three sessions, allowing a more participation from the patient. This ability to 'disrupt' the traditional pattern of GP consultations is important in an era where there is recognition in New Zealand and other OECD countries about the need to respond to the changing context of primary care, particularly in relation to long term conditions including common mental health problems [42].

The choice of 4 points for a minimal clinically important difference on the K10 measure was selected on the basis of past work [9] . Subsequent research suggests a minimum clinically important difference of around 7 points (measured in younger people accessing services [43]. In retrospect, the selection of a smaller difference to detect for the sample size calculation does not

affect the interpretation of results as the current study would have had more than 80% power to detect this revised larger difference between study groups. The original sample size calculation also indicated that full recruitment would have achieved 80% power to detect a difference of 3.2 points on the HADS scale: this was a slightly bigger difference than the minimal clinically important difference cited in the literature [44].

We also examined the impact of analytical decisions on our primary outcome, particularly sensitivity analyses examining the potential impact of participants with no post-baseline data (excluded from the main analysis) on the reported intervention effect. There was more loss-to-follow-up observed in the UBI group than in the PAU group. These sensitivity analyses showed relatively little impact on our estimates under several sets of assumptions (Supplementary Methods and Results).

Limitations

The difficulties in recruiting a sufficient sample size meant we were unable to establish benefit or rule out substantial inferiority of UBI compared to PAU. The main challenges of recruitment for trials in mental health have been described [45-47]. The current study contained specific additional challenges as outlined below.

Firstly, our recruitment was limited by specific entry criteria. We would have preferred to include all adults aged 18-65 with K10's exceeding 35, but our partner PHO was required to limit access to services to clients within the targeted access criteria. This reduced our ability to recruit our planned sample size.

This meant we did not meet our planned sample size target despite energetic problem-solving over a 3 year recruitment period. It also meant that many GPs were not able to recruit any patients (n=60 of the recruited GPs) or were not using the UBI tool until weeks or even months after training. This casts doubt on how well GPs would have adhered to the approach or recalled the principles, potentially affecting the quality of the intervention delivered.

Secondly, in this New Zealand context, the GPs in the PAU group had access to a sophisticated range of therapy options which included providing extended consultations themselves, as well as

referring patients to psychological therapies such as counselling or CBT delivered by clinical psychologists (Dowell 2009). In addition, during the course of the study there were significant changes to the way in which the external psychological services were delivered in our local PHO, with therapists (mental health professionals) being placed within practices rather than at a central location making it easier for in-house referral. Thus the results may not generalise to settings where these additional therapies are unavailable in day-to-day practice.

These changes made the task of demonstrating non-inferiority more challenging. UBI is consistent with the contemporary primary care stepped care approach that tailors interventions to symptom severity and response to treatment [48]. The intervention tool (UBI) used in this study was developed for sub-threshold mental health syndromes, but was, in practice, applied to moderate-to-severe problems, due to demand from GPs who said they needed higher thresholds in order to be able to recruit patients. In the New Zealand context it appears those needing mental health interventions in primary care have more severe problems than the tool was intended for. The intervention may have performed relatively better than PAU if applied to a mild-to-moderate group, but this would need further research to ascertain. The moderate-to-severe group are likely to require longer, more intensive interventions for it to make a difference.

Given the known efficacy of the PAU intervention in this setting [9], the results also attest to the success of the PAU options rather than a specific failing of the intervention. Clinicians who participated in this study might be expected to be those who were motivated and skilled in supporting patients with mental health problems. It is unclear in this case the extent to which the GPs in the UBI treatment arm were adhering to the structured approach outlined in the treatment manual. Fidelity and adherence to training for psychological intervention has been subject to commentary in the literature [49, 50] and it is unclear as to the extent to which UBI GPs were able to adhere to the structured manual.

The analyses presented here examined several arising issues that were not planned for at the start of the study. Firstly, there were imbalances on some demographic variables (gender and age group) between the two study arms. While this is sub-optimal, the analysis of primary and secondary outcomes adjusted for these and other sociodemographic factors, which means that these imbalances should be accounted for in the results.

Conclusion

In this study both the PAU and UBI groups showed improvement in clinical outcome, despite UBI failing to demonstrate superiority or conclusive non-inferiority compared to PAU. This leaves open the question of whether this style of intervention may have potential value in a primary care setting, or whether some elements of this style of intervention are already being applied in practice by some clinicians. Either way, our results did not show that the UBI added value to usual care with patients with moderate-to-severe symptoms.

An ultra-brief approach such as UBI may add value if restricted to patients with mild mental health problems, as part of a suite of options, with different levels of intensity available to GPs in the primary care setting.

There is a significant need for further research into these issues, given the recognition of mental health problems at a community level [6, 51] and the challenge of providing access to psychological therapy in an effective and cost-effective way [52, 53].

Figure Legends

Figure 1. Study flowchart of patient participation.

Figure 2. Mean K10 score (95% CI) at baseline and follow up for UBI and PAU study arms.

Figure 3. Mean total HADS score (95% CI) at baseline and follow up for UBI and PAU study arms.

Abbreviations

UBI: Ultra-brief intervention; PAU: Practice as Usual; GP: General Practitioner; PHO: Primary Health Organisation; K10: Kessler Psychological Distress Scale; HADS: Hospital Anxiety and Depression Scale; WSAS: Work and Social Adjustment Scale; NZDep2006: New Zealand index of individual socioeconomic deprivation.

Authors' contributions

All authors contributed to the study design and study protocol. FM and SC are co-principal investigators. SC conceived the study, obtained initial funding, and contributed to the

development of the intervention. FM and RT obtained co-funding. FM largely developed the intervention, led GP training and PHO liaison. AD contributed to the intervention design and GP training. JS contributed to the study design and designed and conducted the analysis. JS, FM, AD and SC jointly interpreted the results. RT contributed as research assistant, assisted with practice recruitment and GP training, led the patient recruitment, data collection, processing and project management in the latter stages. All authors contributed to and approved the final manuscript.

Funding

Funding for this study was received from the local partner PHO, Compass Health, The Oakley Mental Health Research Foundation, Wellington Medical Research Foundation, and the University of Otago Research Fund. Funding was also received from the Health Research Council of New Zealand for the initial feasibility study, and from Oakley Mental Health Research Foundation and the University of Otago Research Fund for the Māori adaptation. The funding bodies had no role in the study design, collection, analysis and interpretation of data or in the writing of the manuscript.

Competing Interests

None

Acknowledgements

Thanks to Gabrielle Jenkin, who managed the project in the early phases; Felicity Goodyear Smith, Simon Hatcher and Sarah Gordon, who contributed to the study design in the early phases; Amy Munstermann who facilitated liaison with Compass Health; Brigitte Lane, who recruited practices and led the data collection for the first year and Denise Steers who contributed as a research assistant.

Data sharing

Individual-level patient data are not available to other researchers as participants were not asked for consent to share their data. The study protocol (including statistical analysis plan) is available at [27] (DOI:10.1186/s13063-015-0778-y). The code used to conduct the statistical analysis is available from the second author on request (james.stanley@otago.ac.nz).

References

1. Oakley-Browne M, Wells J, Scott K. Te Rau Hingengaro: The New Zealand Mental Health Survey. Wellington: Ministry of Health; 2006.
2. Greenberg PE, Kessler RC, Birnbaum HG, et al. The economic burden of depression in the United States: How did it change between 1990 and 2000? The Journal of Clinical Psychiatry. 2003;64(12):1465-75.
3. Wang PS, Simon G, Kessler RC. The economic burden of depression and the cost-effectiveness of treatment. International Journal of Methods in Psychiatric Research. 2003;12(1):22-33.
4. Horton R. Understanding disease, injury and risk. The Lancet. 2012;380:2053-4.
5. Thornicroft G, Chatterji S, Evans-Lacko S, et al. Undertreatment of people with major depressive disorder in 21 countries. The British Journal of Psychiatry. 2017;210(2):119-24.
6. The MaGPIe Research Group. The nature and prevalence of psychological problems in New Zealand primary health care: a report on Mental Health and General Practice Investigation. 2003. Contract No.: 1171.
7. Layard R, Clark D, Bell S, et al. The depression report; A new deal for depression and anxiety disorders. London School of Economics; 2006.
8. Hollinghurst S, Kessler D, Peters TJ, et al. Opportunity cost of antidepressant prescribing in England: Analysis of routine data. BMJ: British Medical Journal. 2005;330(7498):999-1000.
9. Dowell A, Garrett S, Collings S, et al. Evaluation of the Primary Mental Health Initiatives: summary report 2008. Wellington: Otago University Wellington and Ministry of Health; 2009.
10. The MaGPIe Research Group. The treatment of common mental health problems in general practice. Fam Pract. 2006;23:53-9.
11. Mathieson F, Collings S, Dowell A. Sub-threshold mental health syndromes: Finding an alternative to the medication of unhappiness. Journal of Primary Health Care. 2009;1(1):74-7.
12. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (5th ed.). Washington, DC2013.
13. Wagner H, Burns B, Broadhead W, et al. Minor depresssion in family practice: functional morbidity, co-morbidity, service utilisation and outcomes. Psychological Medicine. 2000;30:1377-90.
14. Collings S, MaGPIe Research Group. Disability and the detection of mental disorder in primary care. Soc Psychiatry Psychiatr Epidemiol. 2005;40(12):994-1002.
15. Kessler R, Zhao S, Blazer D, et al. Prevalence, correlates, and course of minor depression in the national comorbidity survey. Journal of Affective Disorders. 1997;45(1-2):19-30.
16. Sadek N, Bona J. Subsyndromal symptomatic depression: a new concept. Depression and Anxiety. 2000;12:30-9.
17. Dowell A, Garrett S, Collings S, et al. Primary Mental Health Initiatives: Interim Report. Wellington: School of Medicine & Health Sciences, University of Otago, Wellington; 2007.
18. Dew K, Dowell A, McLeod D, et al. "This glorious twilight zone of uncertainty": mental health consultations in general practice in New Zealand. Social Science & Medicine 2005;61(6):1189-200.
19. Mathieson F, Collings S, Dowell A, et al. Collaborative research: A case example of dissemination of CBT in primary care. the Cognitive Behaviour Therapist Vol 6 2013, ArtID e4. 2013;6.
20. Kessler D, Lewis G, Kaur S, et al. Therapist-delivered internet psychotherapy for depression in primary care: A randomised controlled trial. The Lancet. 2009;374(9690):628-34.
21. Jorm AF, Griffiths KM. Population promotion of informal self help strategies for early intervention against depression and anxiety. Psychological Medicine. 2006;36:3-6.

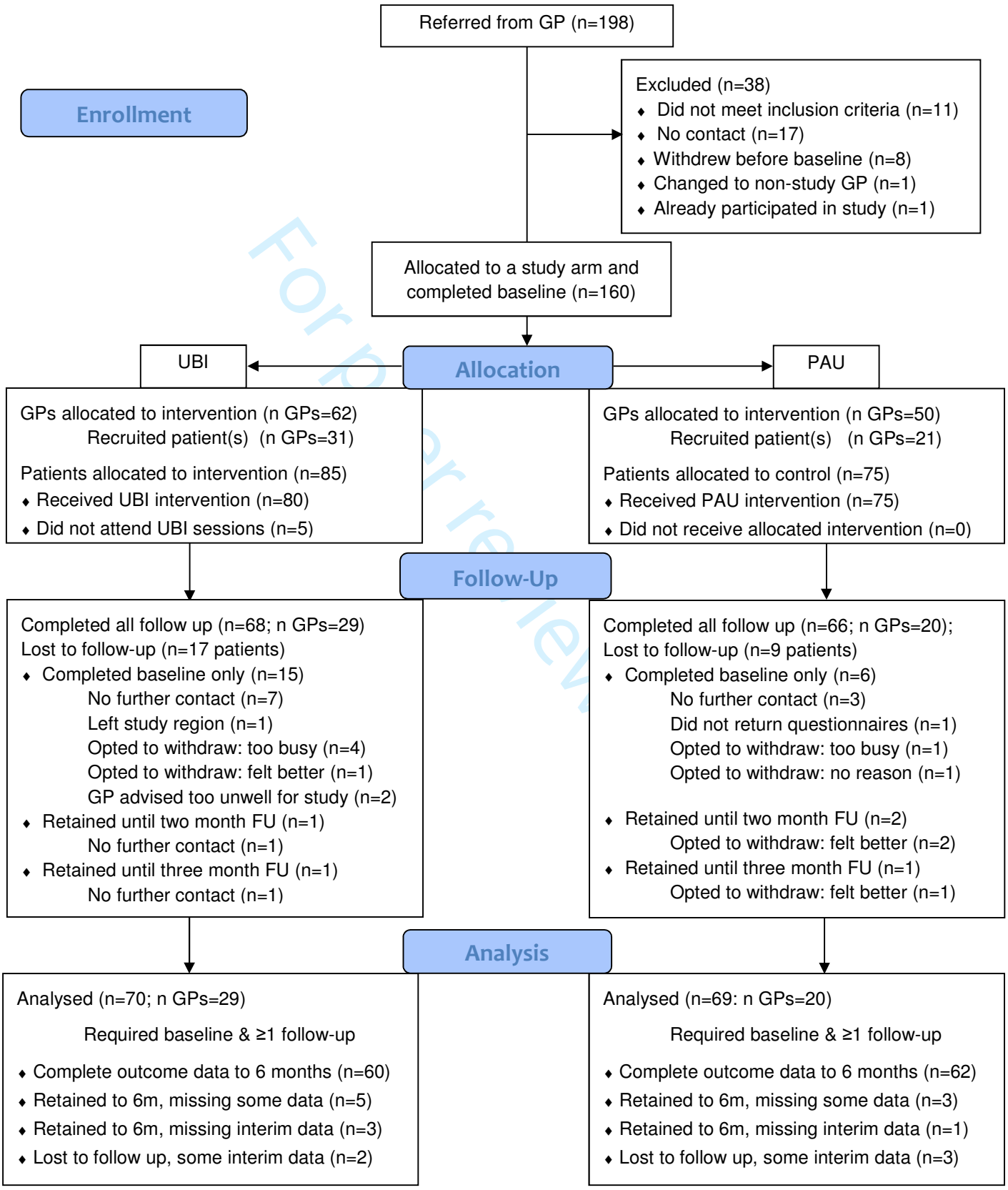
22. Gellatly J, Bower P, Hennessy S, et al. What makes self-help interventions effective in the management of depressive symptoms? Meta-analysis and meta-regression. Psychological Medicine. 2007;37(9):1217-28.
23. Gellatly J, Chisnall L, Seccombe N, et al. @Home eTherapy Service for People with Common Mental Health Problems: An Evaluation. Behavioural and Cognitive Psychotherapy. 2017:1-6.
24. Mathieson F, Mihaere K, Collings S, et al. Maori cultural adaptation of a brief mental health intervention in primary care. Journal of Primary Health Care. 2012;4(3):231-8.
25. Collings S, Mathieson F, Dowell A, et al. Acceptability of a guided self-help mental health intervention in general practice. Fam Pract. 2012;29(1):43-9.
26. Kessler R, Andrews G, Clope LJ, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychological Medicine. 2002;32:959-76.
27. Collings S, Mathieson F, Dowell A, et al. Clinical effectiveness of an ultra-brief intervention for common mental health syndromes in primary care: study protocol for a cluster randomized controlled trial. Trials. 2015;16(260):1-7.
28. Andrews G, Slade T. Interpreting scores on the Kessler Psychological Distress Scale (K10). Australian and New Zealand Journal of Public Health. 2001;25(6):494-7.
29. Ministry of Health. Ethnicity Data Protocols for the Health and Disability Sector. In: Health Mo, editor. Wellington, New Zealand2004.
30. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatrica Scandinavica. 1983;67(6):361-70.
31. Mundt JC, Marks IM, Shear MK, et al. The work and social adjustment scale: a simple measure of impairment in functioning. British Journal of Psychiatry. 2002;180:461-4.
32. Spinhoven P, Ormel J, Sloekers PP, et al. A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects. Psychological Medicine. 1997;27(2):363-170.
33. Bates D, Maechler M, Bolker B, et al. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software. 2015;67(1):1-48.
34. van Buuren S, Groothuis-Oudshoorn K. Mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software. 2011;45(3):1-67.
35. DeSouza CM, Legedza AT, Sankoh AJ. An overview of practical approaches for handling missing data in clinical trials. Journal of Biopharmaceutical Statistics. 2009;19(6):1055-73.
36. Beunckens C, Molenberghs G, Kenward MG. Direct likelihood analysis versus simple forms of imputation for missing data in randomized clinical trials. Clinical Trials. 2005;2(5):379-86.
37. Sullivan TR, White IR, Salter AB, et al. Should multiple imputation be the method of choice for handling missing data in randomized trials? Statistical Methods in Medical Research 2018;27(9):2610-26.
38. Sunderland M, Wong N, Hilvert-Bruce Z, et al. Investigating trajectories of change in psychological distress amongst patients with depression and generalised anxiety disorder treated with internet cognitive behavioural therapy. Behaviour research and therapy. 2012;Jun 30;50(6):374-80.
39. Cape J, Whittington C, Buszewicz M, et al. Brief psychological therapies for anxiety and depression in primary care: meta-analysis and meta-regression. 2010;8(38).
40. Gilbody S, Littlewood E, Hewitt C, et al. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. British Medical Journal. 2015;351(h5627).
41. Andersson G, Carlbring P, Hadjistavropoulos HD. Internet-based cognitive behavior therapy. The science of cognitive behavioral therapy. San Diego, CA: Elsevier Academic Press; US; 2017. p. 531-49.
42. Baird M, Blount A, Brungardt S, et al. Joint principles: integrating behavioral health care into the patient-centered medical home. The Annals of Family Medicine. 2014;12(2):183-5.

43. Rickwood D, Mazzer K, Telford N, et al. Changes in psychological distress and psychosocial functioning in young people visiting headspace centres for mental health problems. The Medical Journal of Australia. 2015;202(10):537-42.
44. Puhan MA, Frey M, Buchi S, et al. The minimal important difference of the hospital anxiety and depression scale in patients with chronic obstructive pulmonary disease. . Health and Quality of Life Outcomes. 2008;6(46).
45. Weisfeld V, English RA, Claiborne AB. Public Engagement and Clinical Trials: New Models and Disruptive Technologies: Workshop Summary.: National Academies Press; 2012.
46. McDonald AM, Knight RC, Campbell MK, et al. What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. Trials. 2006;7(7):1-9.
47. Mason VL, Shaw A, Wiles NJ, et al. GPs' experiences of primary care mental health research: a qualitative study of the barriers to recruitment. Fam Pract. 2007;24(5):518-25.
48. Dowell A, Morris C, Dodds T, et al. Psychological interventions in primary care mental health. In: G. I, editor. Companion to Primary Care Mental Health. London: Radcliffe Publishing; 2012.
49. Bellg AJ, Borrelli B, B. R, et al. Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. Health Psychology. 2004;23(5):443-51.
50. Morton L, Elliott A, Thomas R, et al. Developmental study of treatment fidelity, safety and acceptability of a Symptoms Clinic intervention delivered by General Practitioners to patients with multiple medically unexplained symptoms. Journal of psychosomatic research. 2016;May 31(84):37-43.
51. Whiteford HA, Degenhardt L, Rehm J, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. The Lancet. 2013;Nov 15;382(9904): 1575-86.
52. Gyani A, Shafran R, Layard R, et al. Enhancing recovery rates: lessons from year one of IAPT. Behaviour Research and Therapy. 2013;Sep 30;51(9):597-606.
53. Clark DM. Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: the IAPT experience. International review of psychiatry. 2011;Aug 1;23(4):318-27.

CONSORT

TRANSPARENT REPORTING of TRIALS

CONSORT 2010 Flow Diagram



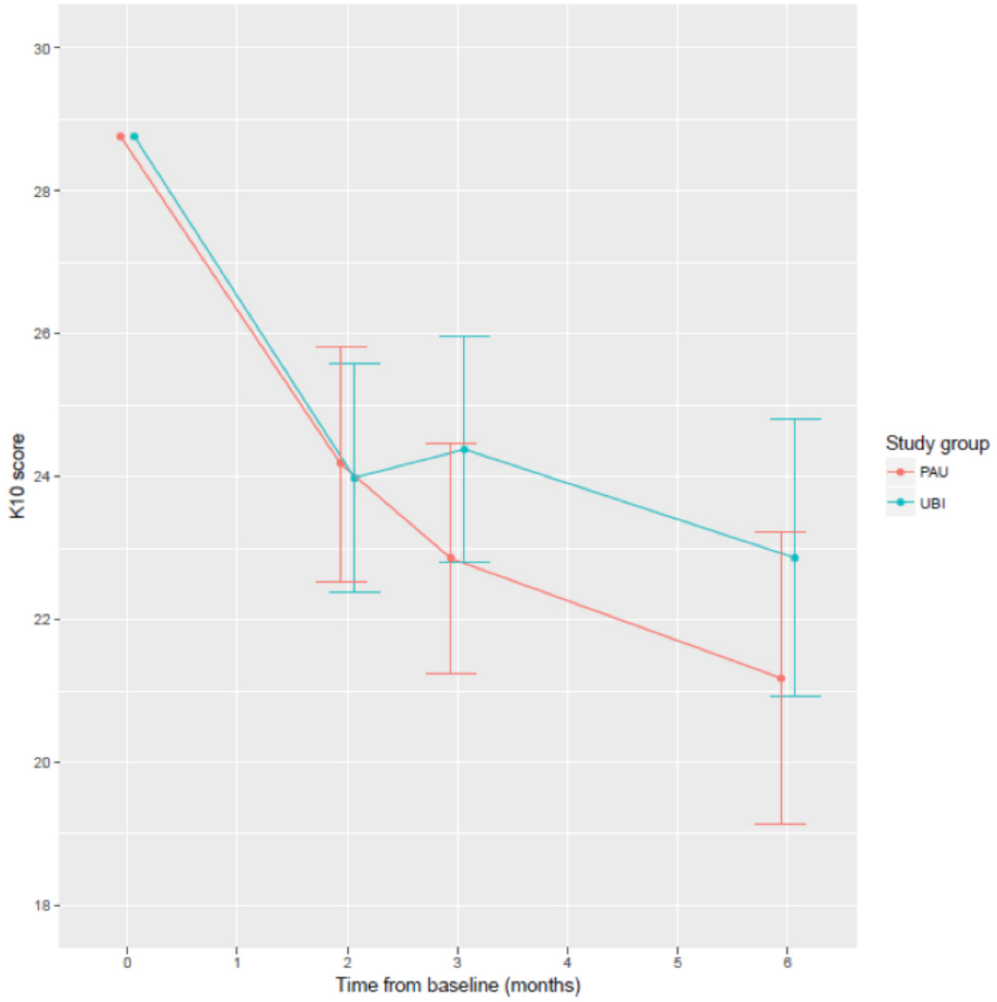


Figure 2. Mean K10 score (95% CI) at baseline and follow up for UBI and PAU study arms.

90x90mm (300 x 300 DPI)

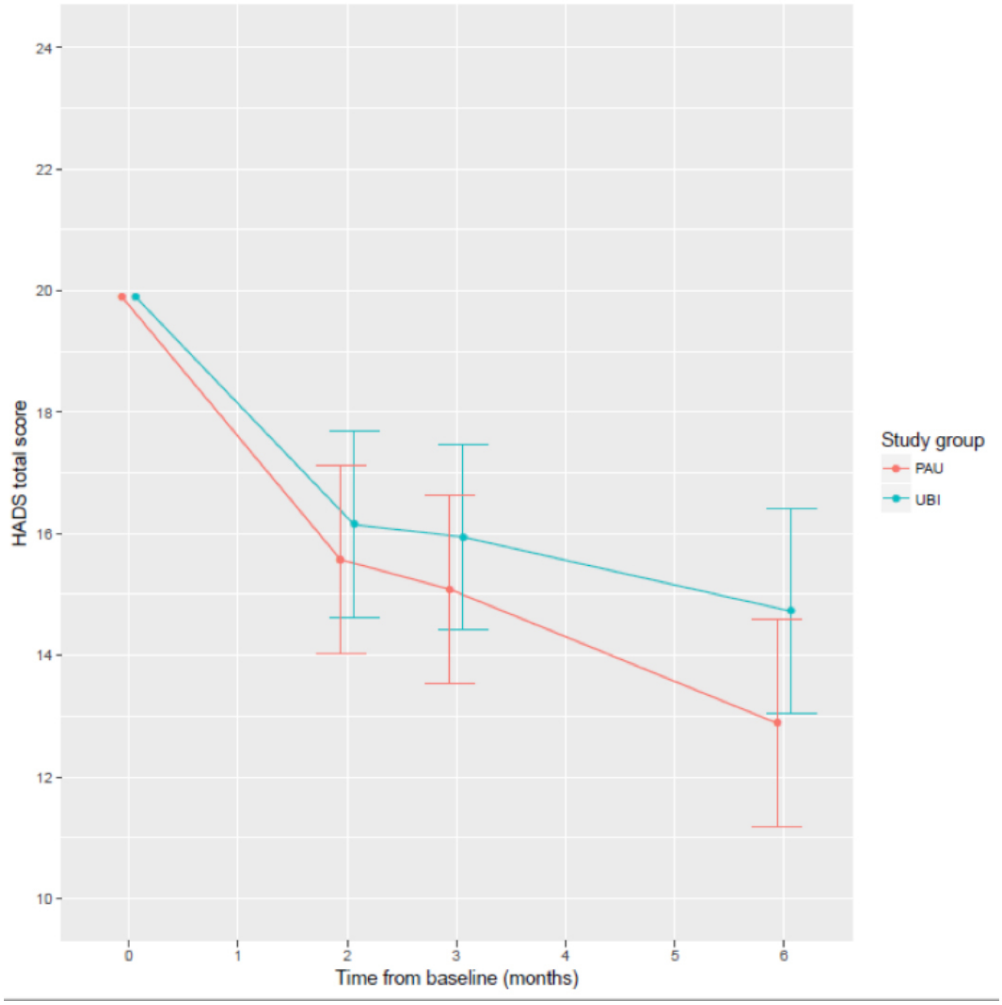
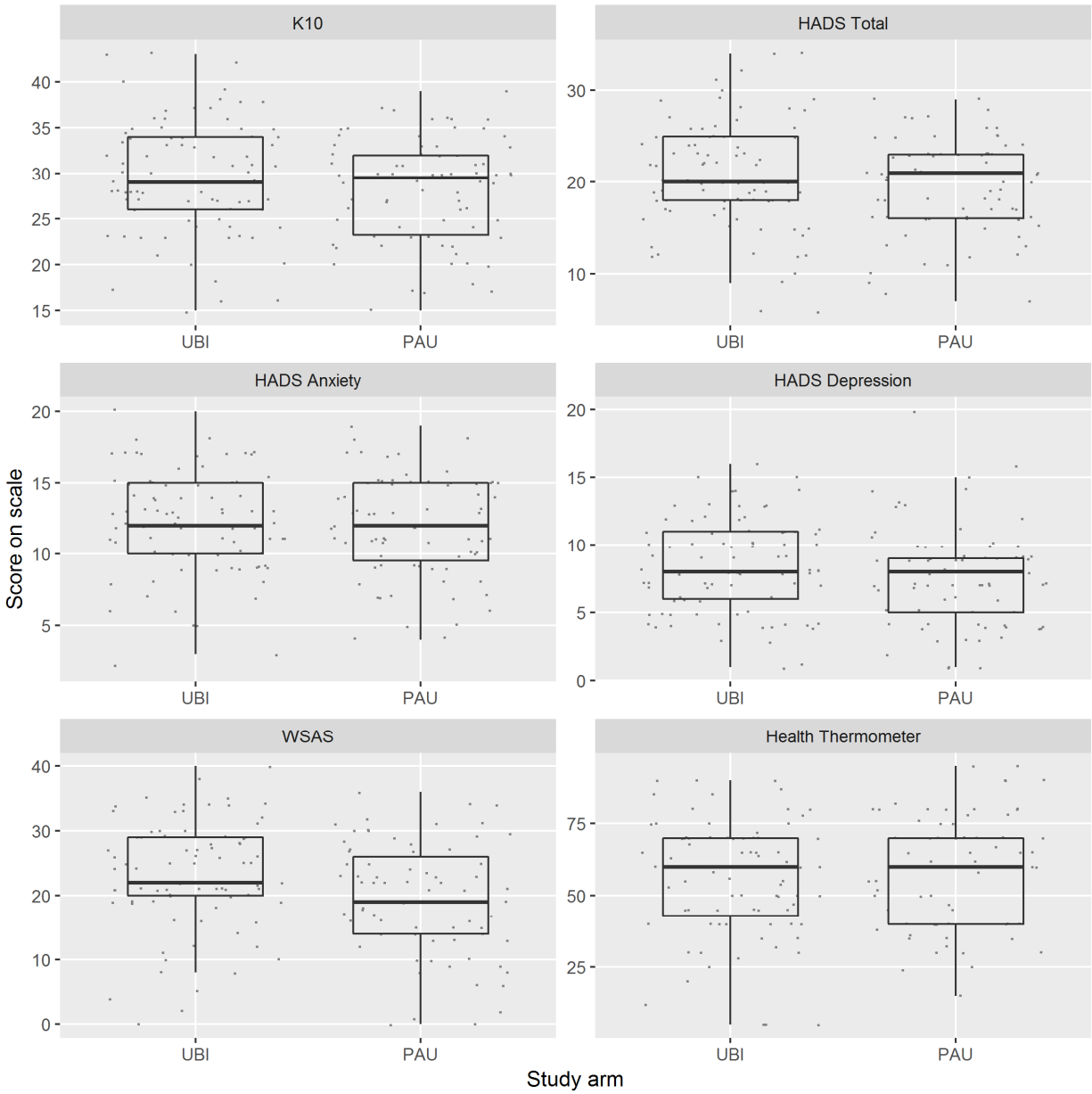


Figure 3. Mean total HADS score (95% CI) at baseline and follow up for UBI and PAU study arms.

90x90mm (300 x 300 DPI)

Supplementary Figure R1. Boxplots of baseline scores for each outcome measure (dots show each individual's score on that measure).



Supplementary Table R1. Number of patients recruited into study by GPs in UBI and PAU study arms.

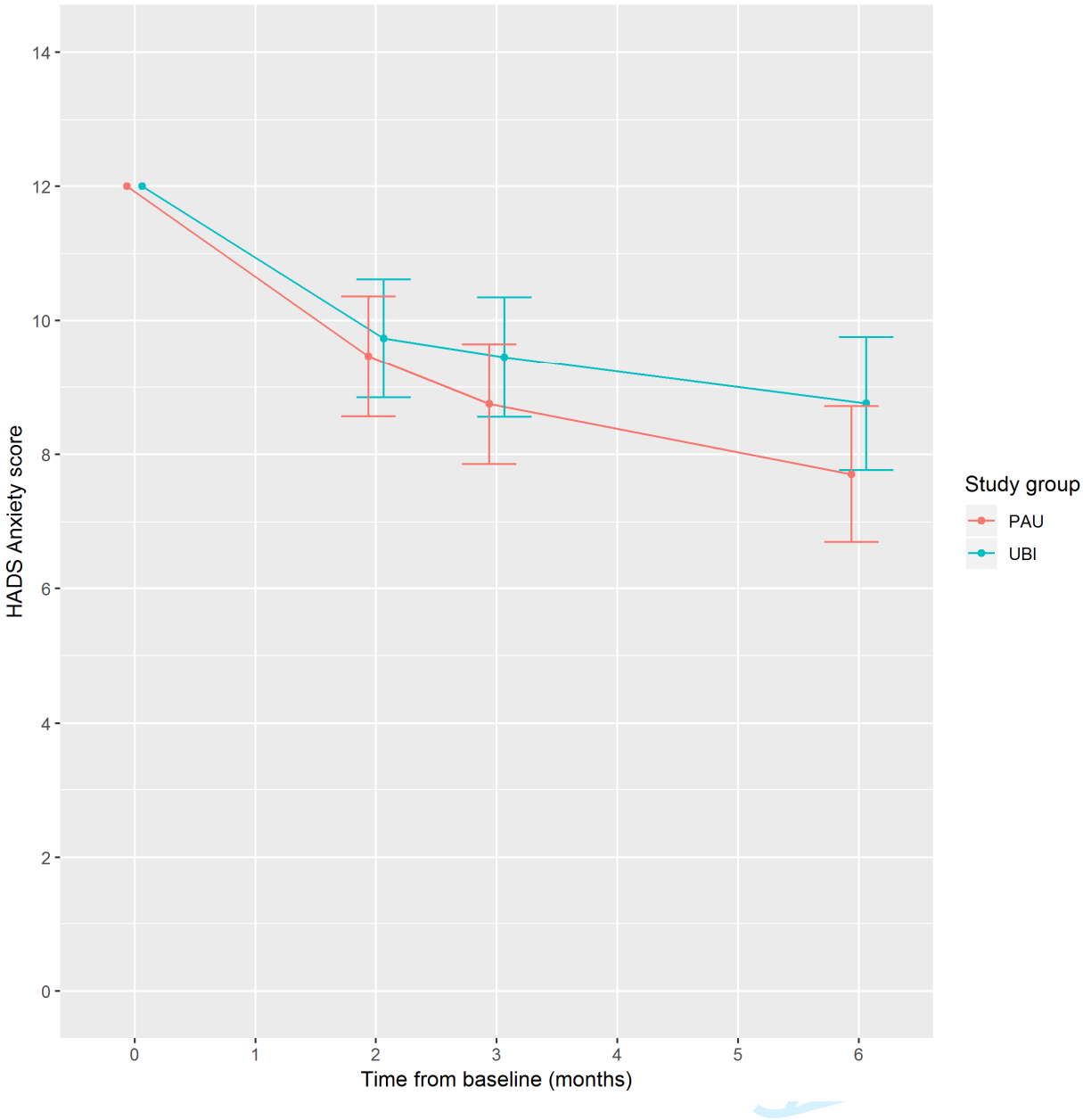
Number of patients recruited by GP	UBI (n GPs*)	PAU (n GPs*)
1	12	8
2	4	2
3	7	5
4	3	0
5	1	2
6	2	0
7	1	0
8	1	1
9	0	2
12	0	1
Total number of GPs	31	21

* Indicates the number of GPs recruiting the stated number of patients (e.g. 12 GPs in the UBI arm recruited one patient each; and five GPs in the PAU arm recruited three patients each).

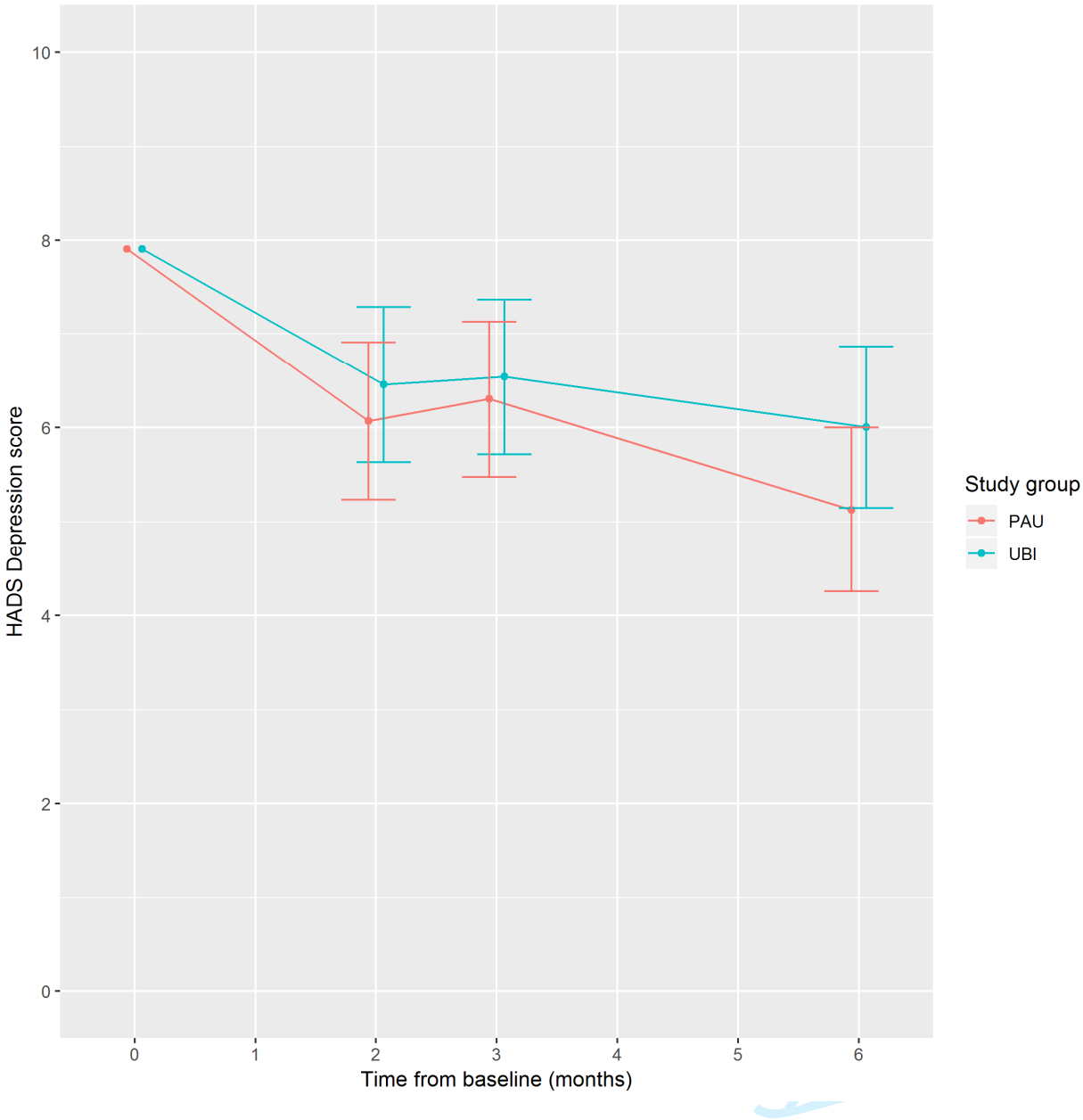
Supplementary Table R2. Mean improvements from baseline to 6 month follow-up for each outcome measure.

Outcome measure	Mean at baseline (both arms)	Mean improvement (95% CI) from baseline to 6 months	
		PAU	UBI
K10	28.8	7.6 (5.5, 9.6)	5.9 (4.0, 7.8)
HADS	19.9	7.0 (5.3, 8.7)	5.2 (3.5, 6.9)
HADS-A	12	4.3 (3.3, 5.3)	3.2 (2.2, 4.2)
HADS-D	7.9	2.8 (1.9, 3.7)	1.9 (1.0, 2.8)
WSAS	21.3	7.7 (5.7, 9.7)	7.2 (5.3, 9.2)
Health Thermometer	57.5	14.0 (9.3, 18.6)	9.0 (4.4, 13.7)

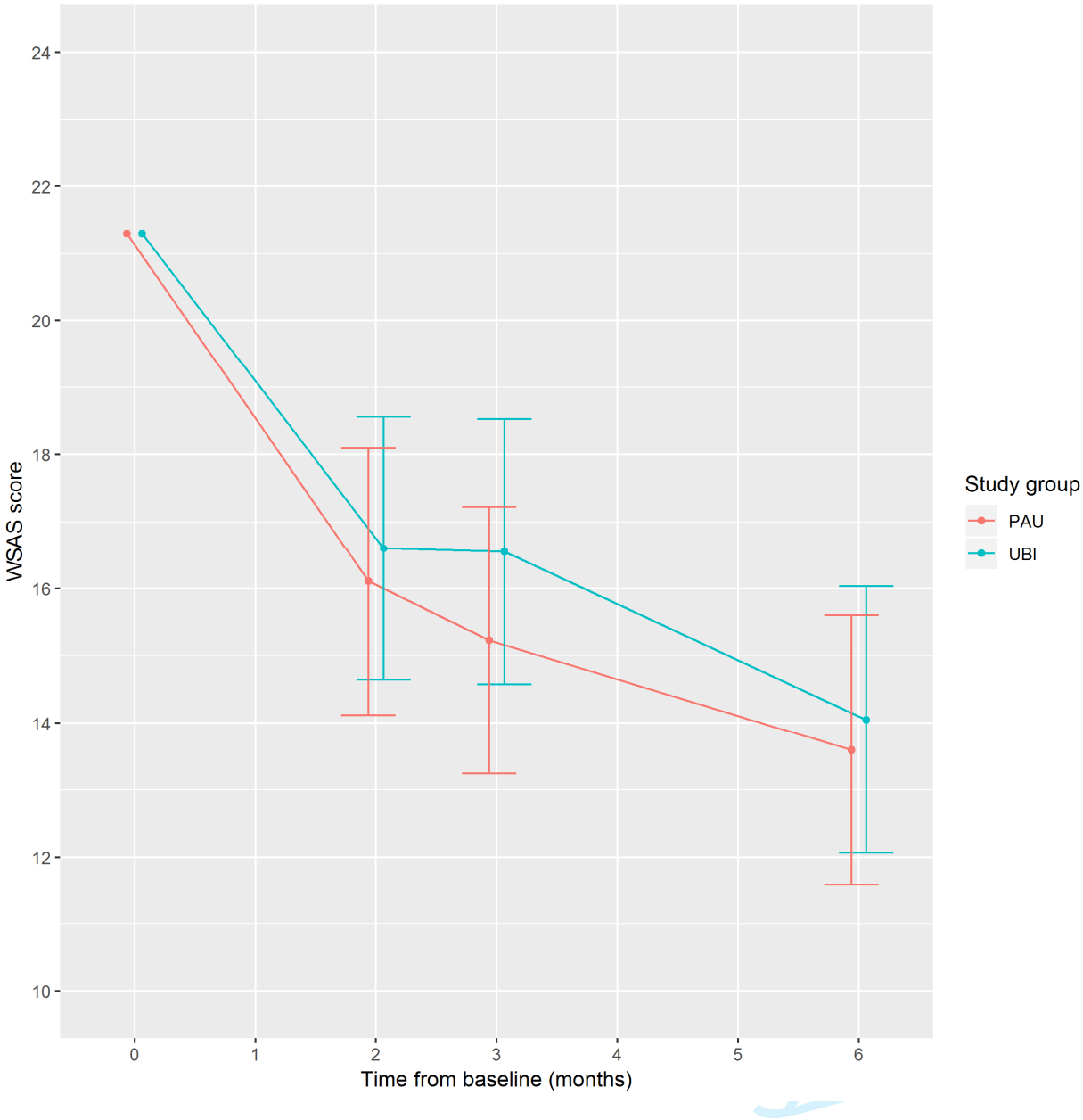
Supplementary Figure R2. Mean HADS Anxiety score (95% CI) at baseline and follow up for UBI and PAU study arms.



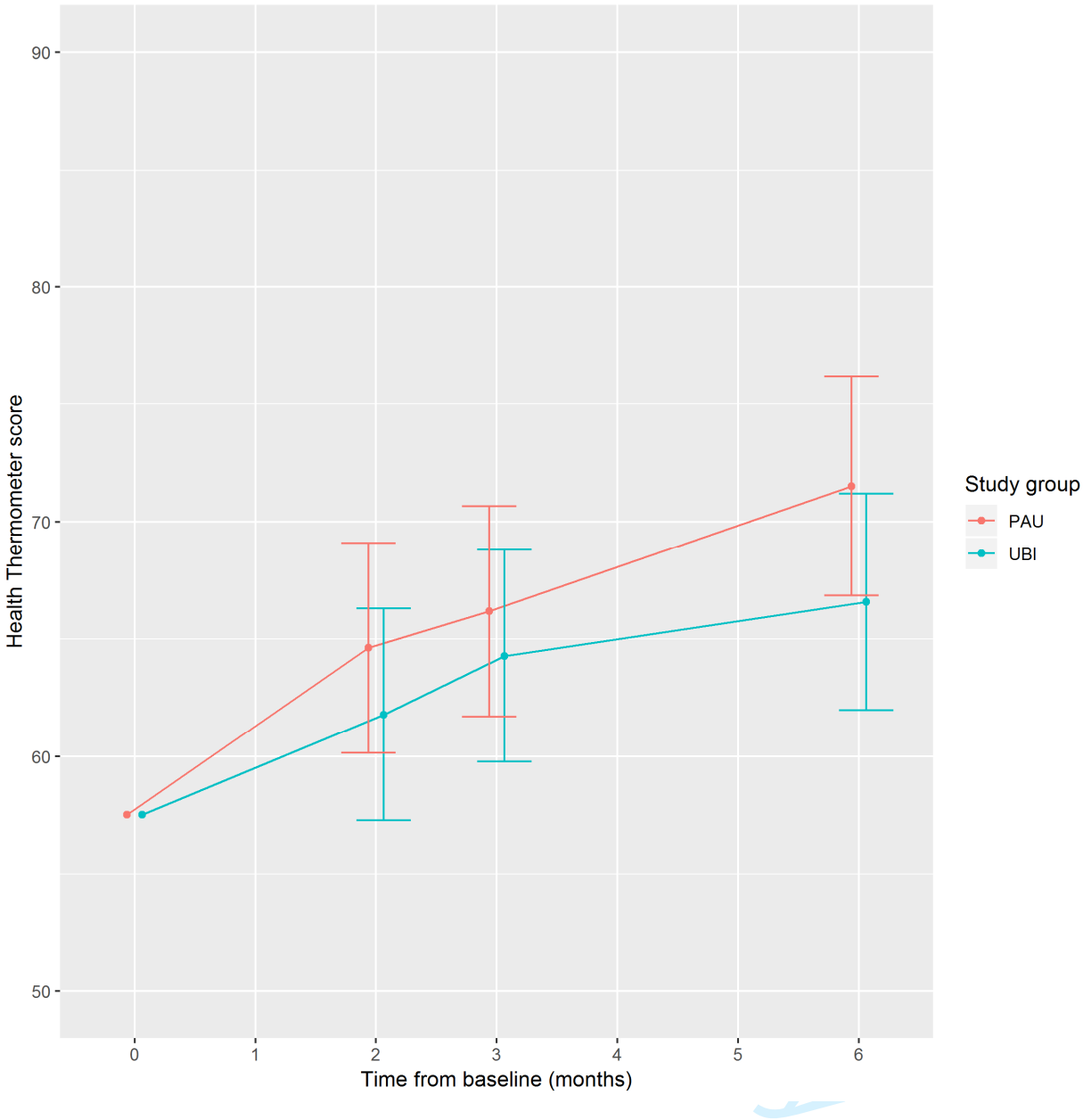
Supplementary Figure R3. Mean HADS Depression score (95% CI) at baseline and follow up for UBI and PAU study arms.



Supplementary Figure R4. Mean WSAS score (95% CI) at baseline and follow up for UBI and PAU study arms.



Supplementary Figure R5. Mean Health Thermometer score (95% CI) at baseline and follow up for UBI and PAU study arms.



Supplementary Table R3. Additional treatment received during UBI trial (from question on 6 month interview)

Type of additional treatment		UBI n (%)	PAU n (%)
Medication status during trial			
	no relevant medication	33 (51%)	34 (52%)
	on medication prior to entering trial	20 (31%)	16 (24%)
	started medication during trial	12 (18%)	16 (24%)
	did not complete question*	20	9
Extended GP consultations (n)			
	0	68 (100%)	46 (71%)
	1-2	0	8 (12%)
	3-5	0	9 (14%)
	6-10	0	2 (3%)
	did not complete question*	17	10
Counselling sessions (n)			
	0	44 (75%)	21 (36%)
	1-2	4 (7%)	13 (22%)
	3-5	2 (3%)	11 (19%)
	6-10	7 (12%)	12 (20%)
	11+	2 (3%)	2 (3%)
	did not complete question*	26	16

* Did not complete 6 month questionnaire and hence no data (UBI n=16; PAU n=9)
Did not answer Meds question at 6 months (UBI: n=4; PAU: n=1)
Did not answer Extended GP question at 6 months (UBI: n=1; PAU: n=1)
Did not answer Counselling question at 6 months (UBI: n=10; PAU: n=7)

Supplementary Methods: Calculation of intra-class correlation coefficients (ICCs) for outcome measures.

ICCs were calculated for each outcome measure in the study to summarise the impact of clustering of outcomes by GPs. These were calculated using simplified mixed linear models with random intercept terms for GPs and no adjustment for covariates. ICCs were calculated in R 3.2.3, using the lme4 package, with their 95% confidence intervals based on 1000 bootstrap resamples calculated using the bootMer() function.

ICCs were also calculated for a scenario where clustering was considered across both the individual GPs (as per the above paragraph) and the practices in which GPs worked. The difference between these two sets of estimates can be considered as the additional impact of clustering of patient responses induced by practices above and beyond clustering induced by GPs. As seen in Supplementary Table R4, there was little impact of this additional clustering on ICCs for the longer health measures (K10 and HADS: minimal difference in ICCs between the two adjustment scenarios) but there appeared to be some additional impact of practice-level clustering for the Work and Social Adjustment Scale (WSAS) and the one-item Health Thermometer.

Supplementary Table R4. Intra-class correlation coefficients (ICCs) for each outcome measure in the study.

	GP clustering only*		GP and Practice clustering**	
Outcome measure	ICC	(95% CI)	ICC	(95% CI)
K10	0.129	(0.045, 0.231)	0.139	(0.006, 0.235)
HADS (total)	0.091	(0.019, 0.189)	0.104	(<0.001, 0.185)
HADS Anxiety	0.098	(0.019, 0.198)	0.106	(<0.001, 0.190)
HADS Depression	0.140	(0.047, 0.250)	0.148	(0.018, 0.233)
WSAS	0.188	(0.081, 0.308)	0.240	(0.076, 0.348)
Health Thermometer	0.088	(0.013, 0.177)	0.135	(0.005, 0.219)

* ICC calculated using only GP-level random effects.

** ICC calculated using random effects for GPs nested within GP practices (joint clustering effect).

Reference for lmer package:

Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Supplementary Table R5. Sociodemographic and clinical characteristics at baseline by intervention arm (UBI or Practice as Usual [PAU]) and follow-up status.

Factor	Level	UBI follow-up (FU) status		PAU follow-up (FU) status	
		Lost to FU	some FU	Lost to FU	some FU
Total	All participants	15 (100%)	70 (100%)	6 (100%)	69 (100%)
Ethnicity	NZE Other	10 (67%)	51 (73%)	5 (83%)	49 (71%)
	Māori	5 (33%)	14 (20%)	1 (17%)	13 (19%)
	Pacific	0 (0%)	4 (6%)	0 (0%)	2 (3%)
	Asian	0 (0%)	1 (1%)	0 (0%)	5 (7%)
Age grp	15-24	11 (73%)	44 (63%)	6 (100%)	31 (45%)
	25-34	2 (13%)	14 (20%)	0 (0%)	15 (22%)
	35-44	1 (7%)	2 (3%)	0 (0%)	13 (19%)
	45-54	0 (0%)	5 (7%)	0 (0%)	6 (9%)
	55+	1 (7%)	5 (7%)	0 (0%)	4 (6%)
Gender	Female	7 (47%)	49 (70%)	3 (50%)	54 (78%)
	Male	8 (53%)	21 (30%)	3 (50%)	15 (22%)
NZiDep	0	3 (20%)	15 (21%)	0 (0%)	11 (16%)
	1	2 (13%)	14 (20%)	1 (17%)	16 (23%)
	2	3 (20%)	12 (17%)	2 (33%)	9 (13%)
	3	0 (0%)	10 (14%)	0 (0%)	10 (14%)
	4	2 (13%)	7 (10%)	1 (17%)	11 (16%)
	5	5 (33%)	12 (17%)	2 (33%)	12 (17%)
Education	At least secondary	15 (100%)	63 (90%)	6 (100%)	65 (94%)
	No secondary	0 (0%)	7 (10%)	0 (0%)	4 (6%)
Outcome scores at baseline		mean (sd)	mean (sd)	mean (sd)	mean (sd)
	K10	28.4 (5.9)	29.8 (6.3)	32.2 (5.3)	27.8 (5.6)
	HADS	20.2 (7.5)	20.7 (5.5)	23.0 (3.0)	19.2 (5.1)
	HADS Anxiety	11.9 (4.9)	12.2 (3.2)	13.2 (2.9)	11.7 (3.5)
	HADS Depression	8.3 (3.2)	8.5 (3.5)	9.8 (3.7)	7.5 (3.6)
	WSAS	21.7 (7.8)	23.3 (8.3)	23.8 (5.2)	19.2 (8.7)
	Health Thermometer*	57.4 (16.5)	55.0 (20.6)	50.5 (17.4)	59.5 (18.7)

* Health Thermometer: Lower scores indicate poorer health state.

Supplementary Results Text 1: Mean difference in K10 primary outcome at 6 months, adjusting only for baseline scores.

The protocol for the primary outcome (K10) analysis only specified that linear mixed model would be adjusted for baseline scores. The results from the primary analysis reported in the main paper were also adjusted for baseline sociodemographic variables (repeated in Supplementary Table R6 below from Table 3).

The analysis of K10 scores at 6 months (adjusted solely for baseline K10 scores) returned a slightly smaller mean difference between groups (poorer mean K10 score in UBI compared to PAU: difference = 1.07, 95% CI -1.67, 3.82).

This supplementary analysis draws on all participants with at least one follow-up observation. All other elements of the statistical model (accounting for clustering by GP and repeat observations for the same participant) are handled as per the main analysis (see Methods of main paper).

Supplementary Table R6. Primary outcome (K10) differences between UBI and PAU study arms at 6 months under different covariate adjustment models.

Analysis	Mean difference in K10 at 6 months (95% CI)
Analysis of all participants with some follow-up (n=139)	
Adjusted for baseline covariates *	1.68 (-1.18, 4.55)
Adjusted for baseline K10 score only**	1.07 (-1.67, 3.82)

* Result as reported in Table 3 of main paper.
** Analysis in line with specifications in protocol paper.

Supplementary Methods and Results: Sensitivity analysis to account for participants with no follow-up data.

The following analyses were implemented following initial peer-review, and were not *a priori* components of the analysis plan. Results from analyses are presented in Supplementary Table XX below, following the description of the methods and results. These sensitivity analyses aimed to consider the impact of complete loss-to-follow-up (participants no post-baseline data) on the primary outcome analysis, using two different frameworks assuming data were missing at random (MAR) or missing not at random (MNAR). A discussion of potential impacts of loss-to-follow-up on study results (attrition bias) is available in Bell et al. (2012) and discussion of missing data mechanisms can be read elsewhere (e.g. Bell et al. (2012); Newgard et al. (2015) and Sullivan et al. (2018)).

References for subsequent section:

Bell ML, Kenward MG, Fairclough DL, Horton NJ. Differential dropout and bias in randomised controlled trials: when it matters and when it may not. BMJ. 2013;346:e8668.

Newgard CD, Lewis RJ. Missing Data: How to Best Account for What Is Not Known. JAMA. 2015;314:940-1.

Sullivan TR, White IR, Salter AB, Ryan P, Lee KJ. Should multiple imputation be the method of choice for handling missing data in randomized trials? Stat Methods Med Res. 2018;27:2610-26.

van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software. 2011;45(3):1-67.

Imputation of outcomes under the Missing at Random (MAR) assumption.

Imputation was implemented using the mice package in R (van Buuren et al., 2011). All primary and secondary outcomes at all follow-up times were included in the imputation model, along with sociodemographic variables at baseline (gender/sex, age group, ethnicity, education, and NZiDep category: see Table 1 of the main paper for details about the specific sub-groups within each of these variables). Imputation was conducted separately for the intervention (UBI) and control (PAU) groups (Sullivan et al., 2018).

A total of 50 imputation datasets were created; each dataset was analysed for the primary outcome following the methods used for the main analysis in the paper (linear mixed model for K10 score at 6 months, adjusted for baseline K10 score and sociodemographic covariates). The estimates from these 50 models were then combined using Rubin's rules to produce the point estimate and 95% confidence interval (which takes into account variability in the effect estimates across all the imputed datasets.)

The intervention effect at 6 months is presented in Supplementary Table R7: under the assumption that the missing data mechanism was MAR (implemented using multiple imputation) there was a mean difference in K10 scores of 1.78 points (95% CI -0.96, 4.51; positive scores indicate better outcomes in the practice as usual [PAU] arm compared to UBI). This was almost identical to the estimates from the linear mixed model reported in Table 3 (repeated in Supplementary Table R7 for reference) which also assumed an MAR mechanism for missing data (conditional on the adjusted baseline variables in that model), but the analysis in the main results only included participants with at least one post-baseline measurement.

Imputation of outcomes under the Missing Not At Random (MNAR) assumption.

Analysis assuming that outcome values were MNAR was repeated under several conditions to explore the potential impact of different types of missing data mechanisms. These analyses all assumed that participants who did not participate in any follow-up did worse than those who participated in at least one follow-up.

In all scenarios, those who were not lost-to-follow-up (i.e. had at least one follow-up measure) kept either their original K10 scores at 6 months, or their imputed values at 6 months (for those with only partial follow-up: using the same imputed datasets as analysed under the MAR assumption). Imputation under MAR principles was considered reasonable for those with at least one follow-up measurement (but no 6-month measurement), as the follow-up measurements were all timed well after the conclusion of the core interventions delivered as part of the trial.

In MNAR Scenario 1: Individuals with no follow-up data were given a K10 score at six months set to 4 points lower than their imputed score.

In MNAR Scenario 2: Individuals with no follow-up data were given the same K10 score at six months that they had at baseline. This is effectively a "last observation carried forward" analysis for those with no follow-up data.

In MNAR Scenario 3: Individuals with no follow-up data were given a K10 score at six months that was 4 points lower than their baseline score.

The outcome analyses were again repeated on the 50 imputed datasets, and the intervention effect results combined across the resulting estimates.

While the effect sizes were slightly different from the main study result (Supplementary Table R7), these assumptions of data being MNAR had relatively minor impact on effect sizes. The most conservative result was under Scenario 1, assuming outcomes for those with no follow-up data were 4 points worse than imputed, returned a mean difference of 2.03, 95% CI -0.63, 4.70.

Note that the confidence intervals with the MNAR sensitivity analyses are likely to be conservative (i.e. not as wide as they should be) because the differences applied from the imputed or baseline values in each scenario were fixed rather than stochastic quantities (i.e. assumes that the applied difference from the imputed or baseline score was always a fixed quantity for all people).

Supplementary Table R7. Estimates of primary outcome (K10) effect size at 6 months under different assumptions of missing outcome profiles in participants with no follow-up data.

Analysis	Mean difference in K10 at 6 months (95% CI)	
Analysis of all participants with some follow-up (n=139)		
Adjusted for baseline covariates (main analysis*)	1.68	(-1.18, 4.55)
Analysis including all randomised participants (n=160)		
Imputed K10 outcome at 6 months (MAR assumption*)	1.78	(-0.96, 4.51)
Imputed K10 outcome at 6 months (MNAR assumptions*)		
1. K10 at 6m set to 4 points worse than imputed	2.03	(-0.63, 4.70)
2. K10 at 6m set to baseline score	1.45	(-0.95, 3.84)
3. K10 at 6m set to 4 point worse than baseline	1.71	(-0.95, 4.37)

* Result as reported in Table 3 of main body of paper.
* MAR (missing at random) and MNAR (missing not at random) assumptions are for the 21 participants lost to follow up (no post-baseline data).

Table 1: CONSORT 2010 checklist of information to include when reporting a cluster randomised trial

Section/Topic	Item No	Standard Checklist item	Extension for cluster designs	Page No *
Title and abstract				
	1a	Identification as a randomised trial in the title	Identification as a cluster randomised trial in the title	Title page
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)[1,2]	See table 2	In abstract
Introduction				
Background and objectives	2a	Scientific background and explanation of rationale	Rationale for using a cluster design	p. 4-5 also p. 6 (methods)
	2b	Specific objectives or hypotheses	Whether objectives pertain to the the cluster level, the individual participant level or both	p. 4
Methods				
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	Definition of cluster and description of how the design features apply to the clusters	p. 4-5
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons		n/a
Participants	4a	Eligibility criteria for participants	Eligibility criteria for clusters	p.4 (for both clusters and participants)
	4b	Settings and locations where the data were collected		p. 4
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	Whether interventions pertain to the cluster level, the individual participant level or both	p. 7
Outcomes	6a	Completely defined pre-	Whether outcome measures	p. 8

		specified primary and secondary outcome measures, including how and when they were assessed	pertain to the cluster level, the individual participant level or both	
	6b	Any changes to trial outcomes after the trial commenced, with reasons		n/a
Sample size	7a	How sample size was determined	Method of calculation, number of clusters(s) (and whether equal or unequal cluster sizes are assumed), cluster size, a coefficient of intracluster correlation (ICC or *k*), and an indication of its uncertainty	p. 8-9
	7b	When applicable, explanation of any interim analyses and stopping guidelines		n/a (no interim analysis was applied)
Randomisation:				
Sequence generation	8a	Method used to generate the random allocation sequence		p. 6
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	Details of stratification or matching if used	p. 6
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	Specification that allocation was based on clusters rather than individuals and whether allocation concealment (if any) was at the cluster level, the individual participant level or both	n/a
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	Replace by 10a, 10b and 10c	
	10a		Who generated the random allocation sequence, who enrolled clusters, and who assigned clusters to interventions	p. 6 (Recruitment and Randomisation sections)

	10b		Mechanism by which individual participants were included in clusters for the purposes of the trial (such as complete enumeration, random sampling)	p. 7 (Recruitment procedures sub-section)
	10c		From whom consent was sought (representatives of the cluster, or individual cluster members, or both), and whether consent was sought before or after randomisation	p. 6 (for GPs as the clusters) and p.7 (consent for the patients)
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how		GPs unable to be blinded (p 6) Statistician blinded during analysis (p. 9) Research assistant unable to be blinded
	11b	If relevant, description of the similarity of interventions		n/a
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	How clustering was taken into account	p.9-10 Analysis and clustering noted on p. 9
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses		p. 10
Results				
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	For each group, the numbers of clusters that were randomly assigned, received intended treatment, and were analysed for the primary outcome	Clusters (GPs) noted on p. 11, additional detail in Supplementary Table R1. Individual patients noted on p. 11, flowchart in

				Figure 1 (including who was covered in analysis)
	13b	For each group, losses and exclusions after randomisation, together with reasons	For each group, losses and exclusions for both clusters and individual cluster members	Clusters (GPs) covered on p.11 (no losses or exclusions, other than zero-recruitment which is covered in Supplementary Table R1) Patients covered in Figure 1
Recruitment	14a	Dates defining the periods of recruitment and follow-up		p. 5
	14b	Why the trial ended or was stopped		p. 5
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	Baseline characteristics for the individual and cluster levels as applicable for each group	Individual level characteristics reported in Table 1.
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	For each group, number of clusters included in each analysis	Analysis by original assigned groups (methods, p. 9) Number of participants for each analysis: Table 2, Table 3 Number of clusters (across all analyses): Supplementary Table R1
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95%	Results at the individual or cluster level as applicable and a coefficient of intracluster correlation (ICC or k) for each	Effect size and precision given in all tables and figures, and for outcomes

		confidence interval)	primary outcome	reported in body of text ICC reported on p 17 for primary outcomes, and Supplementary Table R4 for all outcomes.
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended		n/a (no binary outcomes used in study)
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory		ICCs reported on page 17 (as noted above) Information on additional treatment received presented p 17
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms[3])		n/a
Discussion				
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses		P 18-19 (recruitment not completed to planned sample size) p 20-21 (other limitations)
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	Generalisability to clusters and/or individual participants (as relevant)	p. 20-21
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence		(across discussion)
Other information				

Registration	23	Registration number and name of trial registry	p.4
Protocol	24	Where the full trial protocol can be accessed, if available	p. 4, reference list for detail
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	p. 22-23

** Note: page numbers optional depending on journal requirements*

Table 2: Extension of CONSORT for abstracts[1,2] to reports of cluster randomised trials

Item	Standard Checklist item	Extension for cluster trials
Title	Identification of study as randomised	**Identification of study as cluster randomised**
Trial design	Description of the trial design (e.g. parallel, cluster, non-inferiority)	
Methods		
Participants	Eligibility criteria for participants and the settings where the data were collected	**Eligibility criteria for clusters**
Interventions	Interventions intended for each group	
Objective	Specific objective or hypothesis	**Whether objective or hypothesis pertains to the cluster level, the individual participant level or both**
Outcome	Clearly defined primary outcome for this report	**Whether the primary outcome pertains to the cluster level, the individual participant level or both**
Randomization	How participants were allocated to interventions	**How clusters were allocated to interventions**
Blinding (masking)	Whether or not participants, care givers, and those assessing the outcomes were blinded to group assignment	
Results		
Numbers randomized	Number of participants randomized to each group	**Number of clusters randomized to each group**
Recruitment	Trial status[1]	
Numbers analysed	Number of participants analysed in each group	**Number of clusters analysed in each group**
Outcome	For the primary outcome, a result for each group and the estimated effect size and its precision	**Results at the cluster or individual participant level as applicable for each primary outcome**
Harms	Important adverse events or side effects	
Conclusions	General interpretation of the results	
Trial registration	Registration number and name of trial register	
Funding	Source of funding	

[1] Relevant to Conference Abstracts

REFERENCES

1 Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, et al. CONSORT for reporting randomised trials in journal and conference abstracts. *Lancet* 2008, 371:281-283

2 Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG at al (2008) CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Med* 5(1): e20

3 Ioannidis JP, Evans SJ, Gotzsche PC, O'Neill RT, Altman DG, Schulz K, Moher D. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med* 2004; 141(10):781-788.

BMJ Open

A cluster randomised controlled trial of a guided self-help mental health intervention in primary care

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-023481.R2
Article Type:	Research
Date Submitted by the Author:	16-Jan-2019
Complete List of Authors:	Mathieson, Fiona; University of Otago – Wellington, Department of Psychological Medicine Stanley, James; University of Otago, Wellington, Public Health; University of Otago, Wellington Collings, Catherine (Sunny) ; University of Otago Wellington, Deans Department Tester, Rachel; University of Otago, Wellington, Primary Health Care and General Practice Dowell, Anthony; Wellington School of Medicine and Health Sciences, General Practice
Primary Subject Heading:	Mental health
Secondary Subject Heading:	General practice / Family practice, Mental health
Keywords:	MENTAL HEALTH, PRIMARY CARE, BRIEF INTERVENTIONS

A cluster randomised controlled trial of a guided self-help mental health intervention in primary care

Fiona Mathieson[1]
Email: fiona.mathieson@otago.ac.nz
James Stanley [2]
Email: james.stanley@otago.ac.nz
Sunny Collings[3]
Email: deanseauow@otago.ac.nz
Rachel Tester [4]
Email: rachel.tester@otago.ac.nz
Anthony Dowell[4]
Email: tony.dowell@otago.ac.nz

* Corresponding author: fiona.mathieson@otago.ac.nz Ph. 64 4 9186034, Fax 64 4 3855877
[1]Department of Psychological Medicine, University of Otago, PO Box 7343, Wellington South 6242, New Zealand
[2]Biostatistical Group, Dean's Department, University of Otago, PO Box 7343, Wellington South 6242, New Zealand
[3]Office of the Dean and Head of Campus Te Tari Manutaki, Department of Psychological Medicine, University of Otago, PO Box 7343, Wellington South 6242, New Zealand
[4]Department of Primary Health Care and General Practice, University of Otago, PO Box 7343, Wellington South 6242, New Zealand

Word Count: 5801

Abstract

Objectives: To ascertain whether an ultra-brief intervention improves mental health outcomes for patients in general practice with mild-to-moderate mental health concerns.

Trial design: Two-arm cluster randomised controlled trial.

Methods:

Participants: General practitioners (GPs) were invited based on working in a participating general practice. Patients were eligible to participate if aged 18-65, scored ≤35 on the Kessler-10 (K10) and if meeting local mental health access criteria (based on age, low income, or ethnic group).

Interventions: Intervention arm GPs were trained on the ultra-brief intervention (UBI) approach, with participating patients receiving three structured appointments over five weeks.

GPs randomised to Practice as Usual (PAU) did not receive training, and delivered support following their existing practice approaches.

Outcome Measures: Primary outcome was patient-level K10 score at 6 months post-recruitment.

Randomisation: GP practices were randomised to UBI training or PAU at the start of the study.

Blinding: GPs were not blinded to group assignment.

Results:

Numbers randomised: 62 GPs (recruiting 85 patients) were randomised to UBI, and 50 to PAU (recruiting 75 patients).

Numbers analysed: 31 GPs recruited at least one patient in the UBI arm (70 patients analysed), and 21 GPs recruited at least one patient in the PAU arm (69 patients analysed).

Outcome: K10 scores from an intention-to-treat analysis were similar in UBI and PAU arms, with a wide confidence interval (mean adjusted K10 difference = 1.68 points higher in UBI arm, 95% CI -1.18, 4.55; p=0.255). Secondary outcomes were also similar in the two groups.

Conclusions: The UBI intervention did not lead to better outcomes than practice as usual, though the study had lower than planned power due to poor recruitment. The study results can still contribute to the continuing debate about brief psychological therapy options for primary care and their development.

Trial registration: Australia New Zealand Clinical Trials Registry ACTRN12613000041752

Funding: Compass Health, Oakley Mental Health Research Foundation, Wellington Medical Research Foundation, University of Otago Research Fund

Strengths and limitations

- Pragmatic effectiveness trial of a mental health intervention in primary care.
- Intervention included Maori cultural adaptations.
- Recruitment issues limit strength of results.
- Intervention was applied to more severe mental health presentations that it was developed for.
- GP degree of adherence to the intervention tool is unclear.

Keywords

Mental Health, Primary Care

Introduction

Mental health is major aspect of health and poor mental health is highly prevalent in the general community. Consistent with international findings, just under 40% of the New Zealand (NZ) population had met criteria for a diagnosable mental disorder during their life, and roughly a fifth had experienced a mental disorder in the previous year [1].

There is also considerable international concern about the healthcare burden arising from mental health problems and substance abuse [2-4]), with the World Mental Health Survey (of 21 countries) suggesting that only 41% of people with depression received treatment that met even minimal standards [5].

In NZ, as in other OECD countries, mental health problems are common presentations in primary care. Around one-quarter of primary care patients (26.5% and 29.8% of men and women, respectively), attending their general practice in NZ met criteria for a mild-moderate mental health disorder and an estimated 50-70% of mental health concerns are managed exclusively at the primary care level, since secondary care services have become more targeted towards severe and enduring mental illness in recent years [6] .

Internationally there is a call for psychological therapies to be more widely available in primary care [7], and growing unease about increasing levels of antidepressant medications being prescribed compared with the limited resources available for psychological interventions [8]. However, treatment options at the primary care level are limited, with GPs expressing concerns about gaps in services for patients with mild-moderate mental health presentations and a desire to offer a brief intervention themselves [9]. In NZ, GPs reported that as few as 22% of patients with mild-moderate mental health syndromes receive any formal help [10].

Such patient presentations often comprise sub-threshold syndromes [11, 12], and cases of mild-moderate common mental disorder. These are combinations of problems such as anxiety, depression, substance use and interpersonal problems that do not meet the threshold for disorder in standard diagnostic systems such as DSM-5. Often these arise in the context of

social problems or family or economic stress. In NZ, 36% of general practice attendees report anxiety, depression or substance-use, or a combination of these issues [6]. Such presentations can be associated with significant impairment in functioning and suffering [13, 14], with some going on to develop severe depression [15, 16]. Intervention may be warranted for up to 80% of those affected [10, 13], but referral out of the practice can be problematic due to referral eligibility criteria, waiting times, administrative issues and cost [9, 17, 18].

Increasing knowledge of the burden of mild-moderate disorder led to the development of a platform of Primary Mental Health Initiatives in NZ, which included some increase in access to psychological therapies and extended consultations with GPs. The inclusion criteria for these initiatives, however, mean that only up to 15% of the population can gain access to those services [9].

This service-gap led us to develop a GP delivered ultra-brief intervention (UBI), with development and refinement based on service user feedback [19]. This model has the advantages of avoiding the need for referral on to an expensive professional, such as a psychologist, of being easily accessible to patients, and of potentially building on existing trusted relationships. This fits with the movement towards alternative methods of service delivery for mild to moderate mental health presentations, often termed 'low intensity' interventions. These interventions often include guided self-help, bibliotherapy and computerised delivery of care, with current evidence suggesting that even minimal therapist contact leads to better outcomes than self-help alone [20-23].

UBI was feasibility tested with a group of 16 patients and then adapted for Maori (the indigenous people of New Zealand) and feasibility tested with a group of 9 patients [24, 25]. Based on questionnaire feedback, clinician and patient satisfaction ratings for both feasibility studies were very positive in terms of relevance and acceptability. The psychological well-being of the patients, as measured by the Kessler-10 (K10) [26], was also significantly improved post-intervention (at 3 month follow-up) for both Maori and non-Maori, although there was no control group [24, 25]. Based on these initial findings we designed a cluster randomized controlled trial to measure the effectiveness of UBI.

The aims of the study were to compare patient-level outcomes on (1) mental health state (as measured by K10 scores) at 6 months between UBI and practice as usual (PAU) study arms (primary outcome) and (2) levels of distress (depression and anxiety) and functioning (work,

social and relationship) at 8 weeks and 3 months between UBI and PAU study arms (as secondary outcomes).

Methods

A protocol for this study has been previously published, and includes description of planned analyses [27]. The trial was registered prior to recruitment commencing with the Australia New Zealand Clinical Trials Registry (registration ACTRN12613000041752.)

Design

We used a pragmatic two-arm single blinded, cluster randomised controlled trial of UBI compared with PAU, in a primary care setting. GPs were randomised by practice to exclusively deliver either UBI or PAU to all their recruited patients. GPs were treated as the clusters in the study design (while there was clustering by practice, the GPs were treated as the unit of analysis as practitioner attributes were anticipated to be a higher source of variability in outcomes.) Analysis followed an intention-to-treat approach.

Setting

The study was conducted in general practices in the greater Wellington region, New Zealand. This included practices in both city and semi-rural settings, serving populations from a wide range of socio-economic backgrounds. Recruitment took place between 1/5/2013 and 1/7/2016. The trial ended prior to achieving the final sample size when funding for data collection was exhausted.

Participants

This was a pragmatic trial supported within existing treatment services. GPs were eligible to participate if they were currently working in a practice that was part of the Compass Health Primary Health Organisation (PHO) which covers the greater Wellington region.

Patients were eligible if aged between 18 and 65 and identified by their GP in a routine appointment as experiencing stress or distress. Patients were required to score 35 or less on

the Kessler Psychological Distress Scale (K10) [26, 28] during their initial GP consultation, with no lower cut-off on this score. The present study followed previous study protocols [24, 25] by including scores between 30 and 35 on the K10 as indicative of mild to moderate levels of psychological distress rather than major psychiatric disorder. Individuals taking anti-depressant or other psychiatric medications were eligible to participate in the study.

Patients were excluded if they lacked fluency in English (as the intervention is an English-language based 'talking therapy'); had significant levels of cognitive impairment as determined by the GP; or had reported recent or acute suicidal ideation (i.e., within the previous 2 weeks). Chronic low level suicidality did not exclude an individual from participating. However, GPs were informed of patients who had high scores or suicidality at screening, or for whom referral to appropriate (secondary) mental health services by GPs was indicated, and these patients were not eligible to participate further in the study.

Inclusion criteria were based on the access criteria of a local partner primary health organisation (PHO) to psychological therapies. These criteria were youth (defined as 18-24 years old), or individuals aged 25 years or older with low income, or Māori or Pacific Island heritage.

Recruitment of practices and GPs

Initial recruitment of practices was supported by the partner PHO. GPs were identified using primary health organisation and practice lists. All of the practices contracted under the partner PHO were contacted (N=52) and invited to participate in the study, and an effort was made to contact all of the GPs within these practices by email, telephone or in person. A total of 23 practices initially consented to participate in the study and a further 18 were recruited during the course of the study. Two practices merged and three withdrew (in each case the single participating GP left the practice) leaving a total of 37 practices involved in the study.

Randomisation of practices to study arms

Consenting practices were randomised to provide either UBI or PAU to eligible patients. Randomisation was conducted at the practice level to reduce the risk of contamination if GPs from the same practice were assigned to opposite study arms. To ensure approximately equal numbers of GPs per study arm, randomisation of practices was conducted within five strata, according to the number of participating GPs (one/two/three/four/more than four). An

additional two practices dedicated to youth health that were not part of the partner PHO were included and randomised into each arm of the study (i.e. these two practices formed their own stratum). Practices were entered into the trial following consent from individual participating GPs in that practice. Randomisation of all consenting practices was conducted following this step by the project biostatistician (JS) using a computer-based randomisation following the above stratification profile.

GPs randomised to the UBI study arm completed a single two-hour training session (as previously described [25]). Due to the training nature of the intervention, it was not possible to blind GPs as to their study arm allocation.

Recruitment procedures

GPs identified patients with common mental health problems who might fulfil study criteria during routine appointments. These patients were screened by the GP for eligibility (using the K10), and referred to the study team. A research assistant then contacted potential participating patients, met with them in person where possible to explain the study, confirm eligibility, obtain consent to participate, and collect pre-treatment (baseline) data. Measures were then collected by mail or email at post-treatment (8 weeks, 3 months and 6 months). Patients received compensation (NZ $30 [US$21] vouchers, and entry into a draw for an iPad) following the completion of the final questionnaire, to recompense for time and effort in participating in the study.

Intervention

UBI is a low intensity self-management programme which can be delivered by a GP after a single two-hour training session using a treatment manual based on structured problem solving, motivational interviewing and cognitive behaviour therapy (supported with self-help booklets on relationships, bodily stress, breaking habits and stress management).

Patients who consented and completed the intake data collection (K10 and baseline measurements) received the GP-led intervention in three short, structured face-to-face sessions (one 30 and two 15 minute sessions) over a five to six week period. Relevant booklets were provided to the patient after the first session, to be used in the following session. In New Zealand a stepped care approach to management guides the practitioner

towards using the most appropriate therapy option for the severity of presentation. UBI was designed for mild to moderate presentations and in training GPs were comfortable with the use of the UBI approach for first line management. The study protocol allowed for patients in either study arm to alter their treatment as needed (e.g. access other talking therapies, or commence mental health medications). Patients were blinded as to their study allocation in that patients in PAU practices were not informed that the UBI was offered in practices randomised to deliver UBI. They were simply told that the study was looking at the effectiveness of PAU [27].

Practice as usual

Patients in the PAU study arm received GP support delivered according to their practice as usual (and available existing services). PAU typically consists of supportive counselling in a 15 minute face-to-face consultation, the provision of psychotropic medication, referral to psychological or other counselling options, or referral to relevant community services.

Patient characteristics

Patients are described on the basis of age, gender, prioritised ethnicity and NZiDep, a NZ-developed index [29] of individual-level socioeconomic deprivation.

GPs in practices assigned to the PAU study arm received optional training in the intervention at the end of the study.

Patient and Public Involvement

This study had input from an academic mental health consumer (i.e. an academic who is also a mental health service user and who conducts research from a service user perspective) as part of the research team at the feasibility stage, and designed the intervention based on feedback from a focus group process with potential patient users of the mental health intervention which asked what characteristics such an intervention would need to have. This collaborative process is fully described in [19]. This RCT did not have academic consumer or patient involvement in the recruitment to and conduct of the study and the burden of the intervention was not assessed by the patients. Results of this study will be disseminated by email to GP participants who indicated they wanted them on the consent form.

Outcome measures

The primary outcome measure was the K10 scale [26, 28] score at 6 months (adjusted for score at baseline: see analysis). The K10 is widely used as a clinical outcome measure in Primary Care and General Practice in NZ [9]. A 6 month follow up period was chosen to obtain a sufficient period of assessment following the end of the intervention while at the same time balancing out challenges in patient cohort retention. All analyses were conducted to look at patient-level outcomes.

Secondary outcomes were:

1) Hospital Anxiety and Depression Scale (HADS), which measures the severity of depressive and anxiety symptoms in outpatient hospital settings [30]. Reductions in HADS score indicate reduced anxiety and depression.
2) Comparison of K10 scores by treatment group at 8 weeks and 12 weeks, adjusted for baseline scores (to capture short and medium term effectiveness).
3) Work and Social Adjustment Scale [31], a measure of work, social and relationship functioning) administered at baseline, 8, 12 and 26 weeks.

Outcomes were measured at the same time points in both UBI and PAU groups (baseline, and at 8, 12, and 26 weeks following baseline)

Statistical methods

Sample size and Power analysis

Sample size for the cluster randomised trial was calculated using a simulation method, using standard deviations of patient outcomes from the UBI feasibility study (standard deviation of post-treatment scores = 7.5 [25]). To detect a difference in K10 improvement scores of 6 points in the UBI arm compared with 2 points in the control arm (at 80% power and alpha = 0.05) would require 15 GPs per arm recruiting eight completing patients each on average (n=240 total with complete data). Adjusting for loss to follow-up of 20% gave a recruitment target of ten patients per GP. The simulation settings roughly correspond to an intraclass correlation (ICC) of 0.15 for considering clustering of patient scores by GP (equivalent to the ICC from the feasibility study [25]). Power analysis for the secondary HADS outcome indicated 80% power to detect a difference of 3.2 points between groups (based on a standard

deviation of approximately 6 [32]) assuming a similar ICC for the HADS scale as for the K10 measure (empirical data were not available).

Data Analysis

The statistician was blinded to the intervention or control status of participants (both practices and patients) during conduct of the study and analysis. Results were unblinded once analysis was complete. Data processing and analysis were conducted in R 3.2.3 (R Institute, Vienna) with linear mixed models fit using the lmer package [33] and imputation conducted using the mice package [34].

For the primary outcome, K10 scores at 6 months were compared between the intervention and control groups using mixed linear models (comparing post-intervention scores between groups, adjusting for intake score as a covariate, and treating GP clusters as random slope effects). Analysis was conducted on an intention-to-treat basis according to the study arm for each patient at entry into the study. Analyses were adjusted for all other baseline covariates (age, gender, ethnicity, educational level, and NZiDep). The original protocol stated that analyses would only be adjusted for baseline-values of each score: given some slight imbalance in sociodemographic characteristics it was decided to adjust for other baseline covariates in the main analyses. The originally planned analyses are presented in supplementary materials (overall patterns discussed in the body of the results).

Missing data were handled through the mixed linear models approach to the data, which allows for patients with missing data on the final outcome to be included in analyses, which in effect estimates a final outcome value conditional on the observed data at other follow-up times (i.e. validity being predicated under the assumption that the missing observations are missing at random [MAR], conditional on the observed data [35, 36]). Participants missing all follow-up data were excluded from this main analysis. The null hypothesis for this test was that the K10 scores at 26 weeks (adjusted for baseline score) were not different for the intervention and control groups.

Sensitivity analysis for missing follow-up data in the K10 primary outcome were planned and conducted following completion of the main analysis, and hence were not noted in the trial registration or protocol paper. These analyses covered two scenarios: firstly, an analysis with

multiple imputation of missing outcomes, conditional on observed baseline sociodemographics and baseline outcome data. This analysis hence included participants who only had baseline data recorded (excluded from the main mixed models analysis), and assumes that the unobserved outcome data are missing at random conditional on observed data: that is, that individuals who were missing from all follow-up data collections had the same outcome profile (on average) as participants with similar profiles at baseline [37]. The second sensitivity analysis explored this missing at random assumption: those missing data post-baseline were (i) assumed to have scores at 6 months that were four points worse than their imputed score in the first sensitivity analysis; (ii) assumed to have had no improvement from baseline (last observation carried forward); and (iii) assumed to have had poorer outcomes at six months than at baseline (4 points worse than baseline). Full details of the imputation procedure and sensitivity analyses are presented in the Supplementary material, and results are summarised and discussed in the main body of the results and discussion.

For the secondary analysis, differences in mean scores on the K10 outcome were reported at 8 weeks and 3 months (using the same methods as above, within the mixed linear models framework). Analysis of the HADS and WSAS scores at 8 weeks, 3 months and 6 months utilised the same methods as for the K10 outcome. Analysis of outcomes at 8 weeks and 3 months was not specified on the clinical trials registry, but was noted in the previously published protocol paper [27].

The EQ-5D-3L was noted as a secondary outcome for quality of life in the trial registry. This measure was intended as part of an economic analysis that was not implemented, and no other economic data was collected as part of this study.

Intra-class correlation coefficient (ICC) values were calculated for each outcome measure as a summary of clustering according to GPs. Because our analytical models only accounted for clustering at the level of individual GPs, we also examined ICC values when clustering was considered as a multilevel structure (GPs nested within specific practices). Details of the calculation methods are provided in the Supplementary Materials.

Additional treatments received during the trial (including medication and talking therapies) were analysed by study arm, based on self-report data collected at the 6 month follow-up. This descriptive analysis was not specified in the study protocol.

Confidentiality and data management

Consenting patients had their rights explained along with provision for data confidentiality. Paper and digital copies of the data were secured in locked storage on the premises of the University of Otago, Wellington. The questionnaire data was de-identified and entered into a spreadsheet for subsequent analysis.

Ethics approval

Ethical approval was received from the Health and Disability Ethics Committees (HDEC), Ministry of Health (Northern B Health and Disability ethics committee 12/NTB/2).

Adverse events were not anticipated in this trial, and arrangements were made to feedback clinical information to GPs if deemed necessary (e.g., high K10 scores or concerning self-reported statements about a patient's safety) in the course of data collection.

Results

GP Participants

A total of 41 practices agreed to participate, with a total of 112 individual GPs consenting to take part in the study (n=62 for UBI, and n=50 for PAU). Of these GPs, 31 recruited at least one patient into the study in the UBI arm (from 22 practices), and 21 recruited at least one patient in the PAU arm (from 12 practices). The numbers of GPs recruiting different numbers of patients is shown in Supplementary Table R1.

Patient Participants

Figure 1 summarises the flow of patients into the study and participation in the interventions and follow-up. A total of 198 patients were referred into the study, and 160 met eligibility criteria and completed baseline assessments. The vast majority of these completed at least one post-intervention follow-up (70 / 85 in the UBI arm [82%]; and 69/75 in the PAU group [92%]) and hence contributed to the data analysis. These patients represented 29 GPs (from 21 practices) and 20 GPs (from 12 practices) in the intervention and control arms respectively.

<Insert figure 1 about here>

Baseline data

Baseline sociodemographic characteristics of patients are presented in Table 1 for the two study arms. The two groups were roughly comparable at baseline, with a few more male participants and a slightly younger age profile in the UBI arm, but with a greater representation of females in the study overall.

Table 1. Patient sociodemographic profile by study arm.

Variable	Level	Study Group	
		UBI (Total n = 85)	PAU (Total n = 75)
		n (%)	n (%)
Gender			
	Female*	56 (65.9)	57 (76.0)
	Male	29 (34.1)	18 (24.0)
Age Group			
	15-24	55 (64.7)	37 (49.3)
	25-34	16 (18.8)	15 (20.0)
	35-44	3 (3.5)	13 (17.3)
	45-54	5 (5.9)	6 (8.0)
	55+	6 (7.1)	4 (5.3)
Ethnicity			
	NZE/Other	61 (71.8)	54 (72.0)
	Māori	19 (22.4)	14 (18.7)
	Pacific	4 (4.7)	2 (2.7)
	Asian	1 (1.2)	5 (6.7)
Highest education			
	At least secondary	78 (91.8)	71 (94.7)
	No secondary level	7 (8.2)	4 (5.3)
NZiDep			
	0 (least deprived)	18 (21.2)	11 (14.7)
	1	16 (18.8)	17 (22.7)
	2	15 (17.6)	11 (14.7)
	3	10 (11.8)	10 (13.3)
	4	9 (10.6)	12 (16.0)
	5 (most deprived)	17 (20.0)	14 (18.7)

* Includes one individual self-identifying as Female (transgender)

Mean baseline scores on the outcome measures were also similar between the two groups (Table 2, showing means and standard deviations). Boxplots of the distribution of baseline scores on each outcome scale are given in Supplementary Figure R1.

Table 2. Mean (standard deviation) of baseline scores for outcome measures by study arm

Outcome variable	Study Group	
	UBI (Total n = 85)	PAU (Total n = 75)
	mean (sd)	mean (sd)
K10*	29.5 (6.2)	28.1 (5.7)
HADS – total	20.6 (5.9)	19.5 (5.1)
HADS – anxiety	12.1 (3.6)	11.9 (3.5)
HADS – depression	8.5 (3.5)	7.7 (3.6)
WSAS	23.0 (8.2)	19.6 (8.5)
Health Thermometer**	55.4 (19.9)	58.8 (18.7)

* One patient in PAU group missing baseline value.
** Higher scores on the health thermometer indicate better health.

Health Outcomes at Follow-up

For the K10 primary outcome at 6 months the mean difference for UBI compared to PAU arm favoured the PAU arm (mean adjusted difference = 1.68, 95% CI -1.18 to 4.55; p = 0.255: adjusted for age, gender, ethnicity, educational level, and NZiDep), as shown in Table 3 (where positive differences indicate a better outcome for the PAU than UBI arm) While this result indicated no significant difference in K10 scores between the UBI and PAU arms (see Figure 2), each group had a reasonable improvement in K10 score from baseline (see Supplementary Table R2: for the PAU group mean improvement = 7.6, 95% CI 5.5, 9.6; and for the UBI group mean improvement = 5.9, 95% CI 4.0, 7.8).

Table 3. Mean difference in primary and secondary outcomes (difference in change relative to baseline)

	Mean adjusted difference (UBI minus PAU)*					
	8 weeks		3 months		6 months	
Outcome variable	mean diff (95% CI)	p	mean diff (95% CI)	p	mean diff (95% CI)	p
Primary outcomes**						
K10	-0.19 (-2.55, 2.16)	0.872	1.53 (-0.79, 3.84)	0.203	1.68 (-1.18, 4.55)	0.255
HADS	0.57 (-1.68, 2.82)	0.620	0.86 (-1.38, 3.10)	0.456	1.85 (-0.62, 4.31)	0.149
Secondary outcomes**						
HADS-A	0.27 (-1.02, 1.56)	0.684	0.70 (-0.60, 2.00)	0.296	1.05 (-0.39, 2.50)	0.161
HADS-D	0.39 (-0.82, 1.60)	0.533	0.24 (-0.96, 1.44)	0.701	0.88 (-0.38, 2.14)	0.178
WSAS	0.49 (-2.40, 3.38)	0.740	1.32 (-1.58, 4.22)	0.377	0.45 (-2.47, 3.37)	0.762
Health Thermometer	2.84 (-3.64, 9.31)	0.395	1.90 (-4.59, 8.39)	0.569	4.93 (-1.77, 11.62)	0.156

* Positive differences indicate better improvement in PAU than UBI arm, adjusted for baseline value of score and age, gender, ethnicity, educational level, and NZiDep.

** Number of participants contributing data to each analysis: UBI n = 70, PAU n = 69 (except for K10: PAU n = 68)

<insert Figure 2 about here>

The mean adjusted difference on the HADS measure at 6 months between UBI and PAU measures was 1.85 (95% CI = -0.62, 4.31, p = 0.149; see Table 3), though both groups again showed an improvement in mean score from baseline (Supplementary Table R1). Mean scores at each follow-up time are presented in Figure 3.

<insert Figure 3 about here>

Similarly, for all secondary outcome measures (HADS Anxiety and Depression sub-scales, WSAS, and Health Thermometer), the adjusted difference in outcomes at 6 months showed no significant advantage for either UBI or PAU measures (with relatively broad confidence intervals for these differences: see Table 3.)

Estimates of secondary analyses of outcomes at earlier follow-up times (8 weeks and 3 months) are also presented in Table 3. Differences between UBI and PAU were generally most pronounced at the final follow-up (6 months) compared to the interim follow-ups. Trajectories for mean scores in each group are presented in Supplementary Figure R2, Supplementary Figure R3, Supplementary Figure R4 and Supplementary Figure R5.

Ancillary analyses

Supplementary Table R3 presents information on types of additional treatment received for those who completed the 6-month follow-up assessment (summary not specified in protocol). Similar proportions of completing patients between study arms were either on medication for mental health condition(s) at the beginning on the trial (UBI = 31%; PAU 25%), or started medication during the trial (UBI=18%; PAU=25%). Access to extended GP consultations or counselling sessions was higher for the PAU arm than for UBI (no UBI patient had an extended GP consultation, compared to 29% of PAU patients; and 25% of UBI patients had one or more counselling sessions, compared to 64% of PAU patients.)

Intra-class correlation coefficients (ICCs) for the outcome measures are presented in Supplementary Table R4. For the K10 (ICC = 0.129, 95% CI 0.045 – 0.231) this was relatively close to the ICC values used in planning the sample size for the study. We also

examined clustering effects for GPs as nested within GP practice clusters: this additional complexity (not implemented in our main analytical models) had little impact on ICCs for the K10 or HADS measures, though it did suggest slightly higher ICCs (greater clustering of outcomes than considering GPs alone) for the WSAS and Health Thermometer.

We also conducted three sensitivity analyses for our primary outcome of K10 scores at 6 months. These analyses are described in more detail in the Supplementary Methods and Results.

The first sensitivity analysis used the same linear mixed models analysis as the main reported analysis, but adjusted only for baseline values of the outcome score (as specified in the original protocol: no adjustment for other baseline covariates). This returned a slightly smaller mean difference between study arms (again with a poorer mean K10 score in UBI compared to PAU: difference = 1.07, 95% CI -1.67, 3.82; p=0.447) but does not control for the covariate imbalance seen in recruited participants (as shown in Table 1).

The second and third sensitivity analyses both aimed to consider the impact of loss-to-follow-up on the primary outcome analysis, assuming data were missing at random (MAR) or missing not at random (MNAR). Full details of implementation are in the Supplementary Methods. Both analyses include all randomised participants. An initial table gives the baseline covariates for those with and without follow-up in the PAU and UBI groups (Supplementary Table R5).

The analysis of outcomes under an MAR assumption (including all randomised participants) was almost identical to the main results (Supplementary Table R6). Analyses of outcomes under MNAR assumptions were also not substantively different from the main results (Supplementary Table R7): the most conservative result returned a mean difference of 2.03 points on the K10 (95% CI -0.63, 4.70: Scenario 1 in Supplementary Table R7) which was slightly bigger than the mean difference seen in the main results (1.68 points, as per Table 3).

Discussion

The brief psychological treatment (UBI) delivered by GPs in New Zealand in routine practice settings did not lead to better outcomes than practice as usual (PAU) in this pragmatic efficacy trial, with the point estimate for the primary outcome favouring PAU over UBI.

UBI appeared to be slightly less effective than PAU in reducing distress as measured by the K10 (though the difference was not statistically significant). The K10 was originally introduced as an assessment measure of psychological distress, but has also been used to track change in mental health status following intervention [38]. There were no significant differences in the secondary measures either.

We were unable to achieve full recruitment to match the pre-determined sample size: the study recruited 160 eligible participants across both study arms, against our target of 240 participants with complete data. As such, we were unable to rule out non-inferiority of the intervention (UBI) compared to PAU in reducing the disability and distress associated with mild to moderate mental health problems: the bounds of the confidence intervals for the two main outcomes (K10 and HADS measures) included sizable-magnitude better outcomes for PAU over UBI (e.g. the upper bound for the K10 was a 4.55 point advantage for PAU).

Both UBI and PAU arms showed improvement in clinical outcome over the 6 month course of the study. These findings are in keeping with other work which demonstrates clinical effectiveness of brief psychological interventions in primary care settings [39].

These results suggest that GPs in both arms were achieving clinical benefit. We cannot rule out that UBI performs slightly worse than PAU, but our results are inconclusive due to our reduced sample size. For the last 10-20 years in many OECD jurisdictions there has been a focus on improving mental health care provision in primary care settings. In New Zealand this has taken the form of the introduction of locally based primary mental health initiatives, which have increased access to psychological services and provided opportunity for increased engagement (and remuneration) by General Practitioners to undertake mental health consultation work [9]. These opportunities were available to the PAU, and may partially explain the relative success of this 'control' arm in the study.

Strengths of this study

We consider the results of this trial a useful addition to the literature for two reasons. Firstly they describe the introduction of potentially useful adjuncts to existing therapy approaches in primary care in a randomised controlled setting, and secondly the 'negative results' raise questions about the challenges of conducting pragmatic trials of psychological interventions in primary care and also about the nature and effectiveness of PAU treatments. Feedback received from GPs during the training sessions suggested that elements of the UBI such as

active listening, goal-setting; making a specific plan and following up on it are already used in routine practice. UBI had previously been piloted and shown to be both feasible and acceptable to both clinicians and patients in a general practice setting [25]. It was also able to be adapted in a culturally responsive way [24]. During the course of the trial and following its completion there has been significant interest expressed by both patients and GPs in obtaining copies of the booklets and using elements of the UBI approach in routine consultations. Verbal feedback suggests that GPs particularly liked the helpful/unhelpful behaviour chart which was used to discuss how problems were maintained, the explicit linking of emotional responses to physical symptoms and the use of commitment and capability rulers (a motivational interviewing strategy).

There is an active debate about the optimal balance of intervention components for the management of common mental health problems, with an increasingly varied range of options available. Patients potentially have access to traditional face to face intervention with a therapist, access to materials available on the internet, and further access to rapidly developing telemedicine and virtual consultation options [40, 41]. Our study shows that over the course of the trial, patients and GPs were able to adapt the standard pattern of the GP consultation to a series of three sessions, allowing a more participation from the patient. This ability to 'disrupt' the traditional pattern of GP consultations is important in an era where there is recognition in New Zealand and other OECD countries about the need to respond to the changing context of primary care, particularly in relation to long term conditions including common mental health problems [42].

The choice of 4 points for a minimal clinically important difference on the K10 measure was selected on the basis of past work [9] . Subsequent research suggests a minimum clinically important difference of around 7 points (measured in younger people accessing services [43]. In retrospect, the selection of a smaller difference to detect for the sample size calculation does not affect the interpretation of results as the current study would have had more than 80% power to detect this revised larger difference between study groups. The original sample size calculation also indicated that full recruitment would have achieved 80% power to detect a difference of 3.2 points on the HADS scale: this was a slightly bigger difference than the minimal clinically important difference cited in the literature [44].

We also examined the impact of analytical decisions on our primary outcome, particularly sensitivity analyses examining the potential impact of participants with no post-baseline data (excluded from the main analysis) on the reported intervention effect. There was more loss-to-follow-up observed in the UBI group than in the PAU group. These sensitivity analyses showed relatively little impact on our estimates under several sets of assumptions (Supplementary Methods and Results).

Limitations

The difficulties in recruiting a sufficient sample size meant we were unable to establish benefit or rule out substantial inferiority of UBI compared to PAU. While we did not meet our recruitment targets, the confidence intervals for our estimates are appropriately wide (reflecting the achieved sample size) and can be taken as valid plausible bounds for the true intervention effect. The main challenges of recruitment for trials in mental health have been described [45-47]. The current study contained specific additional challenges as outlined below.

Firstly, our recruitment was limited by specific entry criteria. We would have preferred to include all adults aged 18-65 with K10's exceeding 35, but our partner PHO was required to limit access to services to clients within the targeted access criteria. This reduced our ability to recruit our planned sample size.

This meant we did not meet our planned sample size target despite energetic problem-solving over a 3 year recruitment period. It also meant that many GPs were not able to recruit any patients (n=60 of the recruited GPs) or were not using the UBI tool until weeks or even months after training. This casts doubt on how well GPs would have adhered to the approach or recalled the principles, potentially affecting the quality of the intervention delivered.

Secondly, in this New Zealand context, the GPs in the PAU group had access to a sophisticated range of therapy options which included providing extended consultations themselves, as well as referring patients to psychological therapies such as counselling or CBT delivered by clinical psychologists (Dowell 2009). This introduces the possibility of post-randomisation bias in the control arm due to differential receipt of these other treatments: however, we did not collect details from patients on receipt of such treatments, and thus could not address this potential bias in our analyses. In addition, during the course of

the study there were significant changes to the way in which the external psychological services were delivered in our local PHO, with therapists (mental health professionals) being placed within practices rather than at a central location making it easier for in-house referral. Thus the results may not generalise to settings where these additional therapies are unavailable in day-to-day practice.

These changes made the task of demonstrating non-inferiority more challenging. UBI is consistent with the contemporary primary care stepped care approach that tailors interventions to symptom severity and response to treatment [48]. The intervention tool (UBI) used in this study was developed for sub-threshold mental health syndromes, but was, in practice, applied to moderate-to-severe problems, due to demand from GPs who said they needed higher thresholds in order to be able to recruit patients. In the New Zealand context it appears those needing mental health interventions in primary care have more severe problems than the tool was intended for. The intervention may have performed relatively better than PAU if applied to a mild-to-moderate group, but this would need further research to ascertain. The moderate-to-severe group are likely to require longer, more intensive interventions for it to make a difference.

Given the known efficacy of the PAU intervention in this setting [9], the results also attest to the success of the PAU options rather than a specific failing of the intervention. We might expect that clinicians who participated in this study would be those who were motivated and skilled in supporting patients with mental health problems. This is a speculative point, as we did not collect this kind of data on clinician experience, which is a limitation of the study and needs to be considered when thinking about the generalisability of the current results to other settings. It is unclear in this case the extent to which the GPs in the UBI treatment arm were adhering to the structured approach outlined in the treatment manual. Fidelity and adherence to training for psychological intervention has been subject to commentary in the literature [49, 50] and it is unclear as to the extent to which UBI GPs were able to adhere to the structured manual.

The analyses presented here examined several arising issues that were not planned for at the start of the study. Firstly, there were imbalances on some demographic variables (gender and age group) between the two study arms. While this is sub-optimal, the analysis of primary and secondary outcomes adjusted for these and other sociodemographic factors, which means that these imbalances should be accounted for in the results.

Conclusion

In this study both the PAU and UBI groups showed improvement in clinical outcome, despite UBI failing to demonstrate superiority or conclusive non-inferiority compared to PAU, though the incomplete recruitment means that the precision of our estimates of treatment effects were wide (95% CIs). This leaves open the question of whether this style of intervention may have potential value in a primary care setting, or whether some elements of this style of intervention are already being applied in practice by some clinicians. Our results did not show conclusive evidence that the UBI added value to usual care with patients with moderate-to-severe symptoms, and we were unable to rule out the possibility that UBI patients may have marginally worse outcomes than the control group. Despite this uncertainty in the outcome, the results provide valuable additional information about the provision of brief psychological interventions in primary care.

An ultra-brief approach such as UBI may add value if restricted to patients with mild mental health problems, as part of a suite of options, with different levels of intensity available to GPs in the primary care setting.

There is a significant need for further research into these issues, given the recognition of mental health problems at a community level [6, 51] and the challenge of providing access to psychological therapy in an effective and cost-effective way [52, 53].

Figure Legends

Figure 1. Study flowchart of patient participation.

Figure 2. Mean K10 score (95% CI) at baseline and follow up for UBI and PAU study arms.

Figure 3. Mean total HADS score (95% CI) at baseline and follow up for UBI and PAU study arms.

Abbreviations

UBI: Ultra-brief intervention; PAU: Practice as Usual; GP: General Practitioner; PHO: Primary Health Organisation; K10: Kessler Psychological Distress Scale; HADS: Hospital Anxiety and Depression Scale; WSAS: Work and Social Adjustment Scale; NZDep2006: New Zealand index of individual socioeconomic deprivation.

Authors' contributions

All authors contributed to the study design and study protocol. FM and SC are co-principal investigators. SC conceived the study, obtained initial funding, and contributed to the development of the intervention. FM and RT obtained co-funding. FM largely developed the intervention, led GP training and PHO liaison. AD contributed to the intervention design and GP training. JS contributed to the study design and designed and conducted the analysis. JS, FM, AD and SC jointly interpreted the results. RT contributed as research assistant, assisted with practice recruitment and GP training, led the patient recruitment, data collection, processing and project management in the latter stages. All authors contributed to and approved the final manuscript.

Funding

Funding for this study was received from the local partner PHO, Compass Health, The Oakley Mental Health Research Foundation, Wellington Medical Research Foundation, and the University of Otago Research Fund. Funding was also received from the Health Research Council of New Zealand for the initial feasibility study, and from Oakley Mental Health Research Foundation and the University of Otago Research Fund for the Māori adaptation. The funding bodies had no role in the study design, collection, analysis and interpretation of data or in the writing of the manuscript.

Competing Interests

None

Acknowledgements

Thanks to Gabrielle Jenkin, who managed the project in the early phases; Felicity Goodyear Smith, Simon Hatcher and Sarah Gordon, who contributed to the study design in the early phases; Amy Munstermann who facilitated liaison with Compass Health; Brigitte Lane, who recruited practices and led the data collection for the first year and Denise Steers who contributed as a research assistant.

Data sharing

Individual-level patient data are not available to other researchers as participants were not asked for consent to share their data. The study protocol (including statistical analysis plan) is available at [27] (DOI:10.1186/s13063-015-0778-y). The code used to conduct the statistical analysis is available from the second author on request (james.stanley@otago.ac.nz).

References

1. Oakley-Browne M, Wells J, Scott K. Te Rau Hingengaro: The New Zealand Mental Health Survey. Wellington: Ministry of Health; 2006.
2. Greenberg PE, Kessler RC, Birnbaum HG, et al. The economic burden of depression in the United States: How did it change between 1990 and 2000? The Journal of Clinical Psychiatry. 2003;64(12):1465-75.
3. Wang PS, Simon G, Kessler RC. The economic burden of depression and the cost-effectiveness of treatment. International Journal of Methods in Psychiatric Research. 2003;12(1):22-33.
4. Horton R. Understanding disease, injury and risk. The Lancet. 2012;380:2053-4.
5. Thornicroft G, Chatterji S, Evans-Lacko S, et al. Undertreatment of people with major depressive disorder in 21 countries. The British Journal of Psychiatry. 2017;210(2):119-24.
6. The MaGPIe Research Group. The nature and prevalence of psychological problems in New Zealand primary health care: a report on Mental Health and General Practice Investigation. 2003. Contract No.: 1171.
7. Layard R, Clark D, Bell S, et al. The depression report; A new deal for depression and anxiety disorders. London School of Economics; 2006.
8. Hollinghurst S, Kessler D, Peters TJ, et al. Opportunity cost of antidepressant prescribing in England: Analysis of routine data. BMJ: British Medical Journal. 2005;330(7498):999-1000.
9. Dowell A, Garrett S, Collings S, et al. Evaluation of the Primary Mental Health Initiatives: summary report 2008. Wellington: Otago University Wellington and Ministry of Health; 2009.
10. The MaGPIe Research Group. The treatment of common mental health problems in general practice. Fam Pract. 2006;23:53-9.
11. Mathieson F, Collings S, Dowell A. Sub-threshold mental health syndromes: Finding an alternative to the medication of unhappiness. Journal of Primary Health Care. 2009;1(1):74-7.
12. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (5th ed.). Washington, DC2013.
13. Wagner H, Burns B, Broadhead W, et al. Minor depresssion in family practice: functional morbidity, co-morbidity, service utilisation and outcomes. Psychological Medicine. 2000;30:1377-90.
14. Collings S, MaGPIe Research Group. Disability and the detection of mental disorder in primary care. Soc Psychiatry Psychiatr Epidemiol. 2005;40(12):994-1002.
15. Kessler R, Zhao S, Blazer D, et al. Prevalence, correlates, and course of minor depression in the national comorbidity survey. Journal of Affective Disorders. 1997;45(1-2):19-30.
16. Sadek N, Bona J. Subsyndromal symptomatic depression: a new concept. Depression and Anxiety. 2000;12:30-9.
17. Dowell A, Garrett S, Collings S, et al. Primary Mental Health Initiatives: Interim Report. Wellington: School of Medicine & Health Sciences, University of Otago, Wellington; 2007.
18. Dew K, Dowell A, McLeod D, et al. "This glorious twilight zone of uncertainty": mental health consultations in general practice in New Zealand. Social Science & Medicine 2005;61(6):1189-200.
19. Mathieson F, Collings S, Dowell A, et al. Collaborative research: A case example of dissemination of CBT in primary care. the Cognitive Behaviour Therapist Vol 6 2013, ArtID e4. 2013;6.
20. Kessler D, Lewis G, Kaur S, et al. Therapist-delivered internet psychotherapy for depression in primary care: A randomised controlled trial. The Lancet. 2009;374(9690):628-34.
21. Jorm AF, Griffiths KM. Population promotion of informal self help strategies for early intervention against depression and anxiety. Psychological Medicine. 2006;36:3-6.

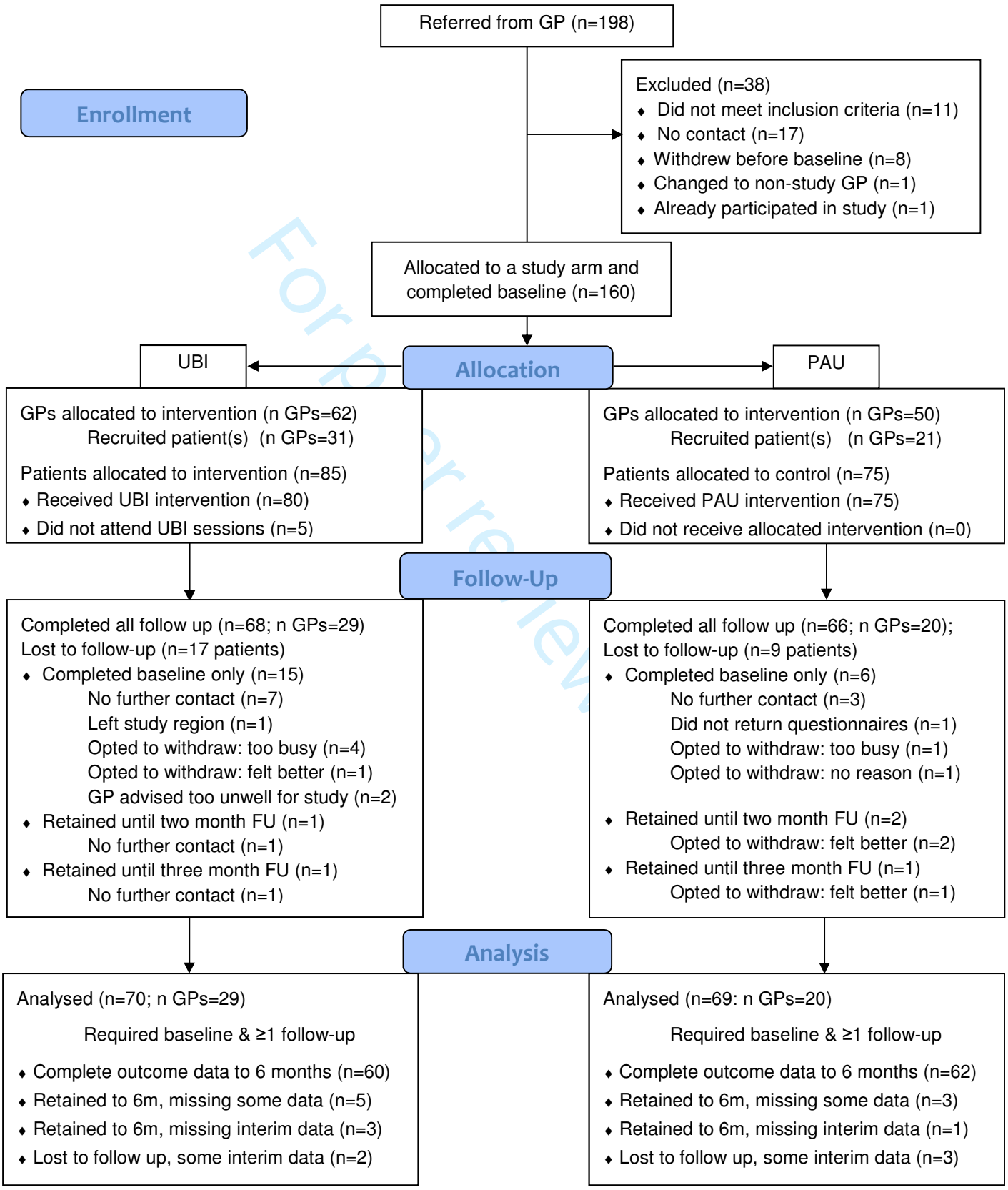
22. Gellatly J, Bower P, Hennessy S, et al. What makes self-help interventions effective in the management of depressive symptoms? Meta-analysis and meta-regression. Psychological Medicine. 2007;37(9):1217-28.
23. Gellatly J, Chisnall L, Seccombe N, et al. @Home eTherapy Service for People with Common Mental Health Problems: An Evaluation. Behavioural and Cognitive Psychotherapy. 2017:1-6.
24. Mathieson F, Mihaere K, Collings S, et al. Maori cultural adaptation of a brief mental health intervention in primary care. Journal of Primary Health Care. 2012;4(3):231-8.
25. Collings S, Mathieson F, Dowell A, et al. Acceptability of a guided self-help mental health intervention in general practice. Fam Pract. 2012;29(1):43-9.
26. Kessler R, Andrews G, Clope LJ, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychological Medicine. 2002;32:959-76.
27. Collings S, Mathieson F, Dowell A, et al. Clinical effectiveness of an ultra-brief intervention for common mental health syndromes in primary care: study protocol for a cluster randomized controlled trial. Trials. 2015;16(260):1-7.
28. Andrews G, Slade T. Interpreting scores on the Kessler Psychological Distress Scale (K10). Australian and New Zealand Journal of Public Health. 2001;25(6):494-7.
29. Ministry of Health. Ethnicity Data Protocols for the Health and Disability Sector. In: Health Mo, editor. Wellington, New Zealand2004.
30. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatrica Scandinavica. 1983;67(6):361-70.
31. Mundt JC, Marks IM, Shear MK, et al. The work and social adjustment scale: a simple measure of impairment in functioning. British Journal of Psychiatry. 2002;180:461-4.
32. Spinhoven P, Ormel J, Sloekers PP, et al. A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects. Psychological Medicine. 1997;27(2):363-170.
33. Bates D, Maechler M, Bolker B, et al. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software. 2015;67(1):1-48.
34. van Buuren S, Groothuis-Oudshoorn K. Mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software. 2011;45(3):1-67.
35. DeSouza CM, Legedza AT, Sankoh AJ. An overview of practical approaches for handling missing data in clinical trials. Journal of Biopharmaceutical Statistics. 2009;19(6):1055-73.
36. Beunckens C, Molenberghs G, Kenward MG. Direct likelihood analysis versus simple forms of imputation for missing data in randomized clinical trials. Clinical Trials. 2005;2(5):379-86.
37. Sullivan TR, White IR, Salter AB, et al. Should multiple imputation be the method of choice for handling missing data in randomized trials? Statistical Methods in Medical Research 2018;27(9):2610-26.
38. Sunderland M, Wong N, Hilvert-Bruce Z, et al. Investigating trajectories of change in psychological distress amongst patients with depression and generalised anxiety disorder treated with internet cognitive behavioural therapy. Behaviour research and therapy. 2012;Jun 30;50(6):374-80.
39. Cape J, Whittington C, Buszewicz M, et al. Brief psychological therapies for anxiety and depression in primary care: meta-analysis and meta-regression. 2010;8(38).
40. Gilbody S, Littlewood E, Hewitt C, et al. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. British Medical Journal. 2015;351(h5627).
41. Andersson G, Carlbring P, Hadjistavropoulos HD. Internet-based cognitive behavior therapy. The science of cognitive behavioral therapy. San Diego, CA: Elsevier Academic Press; US; 2017. p. 531-49.
42. Baird M, Blount A, Brungardt S, et al. Joint principles: integrating behavioral health care into the patient-centered medical home. The Annals of Family Medicine. 2014;12(2):183-5.

43. Rickwood D, Mazzer K, Telford N, et al. Changes in psychological distress and psychosocial functioning in young people visiting headspace centres for mental health problems. The Medical Journal of Australia. 2015;202(10):537-42.
44. Puhan MA, Frey M, Buchi S, et al. The minimal important difference of the hospital anxiety and depression scale in patients with chronic obstructive pulmonary disease. . Health and Quality of Life Outcomes. 2008;6(46).
45. Weisfeld V, English RA, Claiborne AB. Public Engagement and Clinical Trials: New Models and Disruptive Technologies: Workshop Summary.: National Academies Press; 2012.
46. McDonald AM, Knight RC, Campbell MK, et al. What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. Trials. 2006;7(7):1-9.
47. Mason VL, Shaw A, Wiles NJ, et al. GPs' experiences of primary care mental health research: a qualitative study of the barriers to recruitment. Fam Pract. 2007;24(5):518-25.
48. Dowell A, Morris C, Dodds T, et al. Psychological interventions in primary care mental health. In: G. I, editor. Companion to Primary Care Mental Health. London: Radcliffe Publishing; 2012.
49. Bellg AJ, Borrelli B, B. R, et al. Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. Health Psychology. 2004;23(5):443-51.
50. Morton L, Elliott A, Thomas R, et al. Developmental study of treatment fidelity, safety and acceptability of a Symptoms Clinic intervention delivered by General Practitioners to patients with multiple medically unexplained symptoms. Journal of psychosomatic research. 2016;May 31(84):37-43.
51. Whiteford HA, Degenhardt L, Rehm J, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. The Lancet. 2013;Nov 15;382(9904): 1575-86.
52. Gyani A, Shafran R, Layard R, et al. Enhancing recovery rates: lessons from year one of IAPT. Behaviour Research and Therapy. 2013;Sep 30;51(9):597-606.
53. Clark DM. Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: the IAPT experience. International review of psychiatry. 2011;Aug 1;23(4):318-27.

CONSORT

TRANSPARENT REPORTING of TRIALS

CONSORT 2010 Flow Diagram



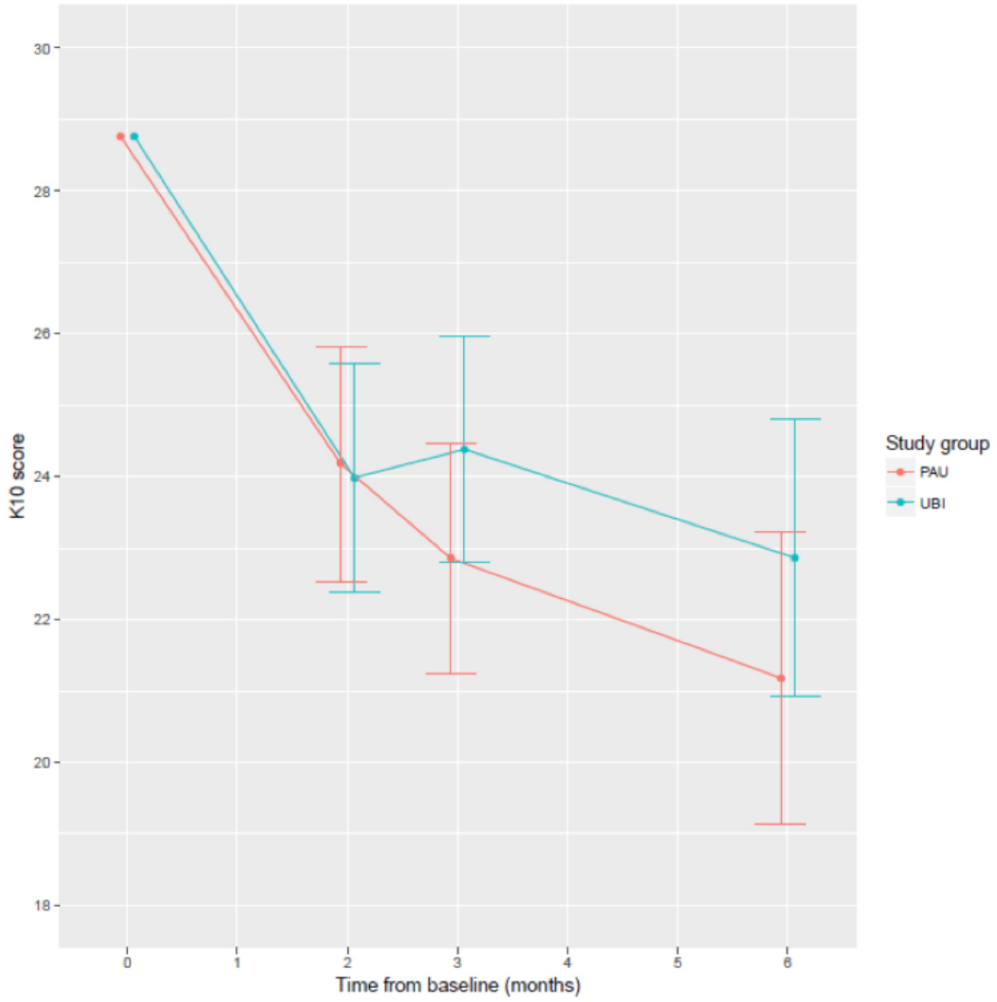


Figure 2. Mean K10 score (95% CI) at baseline and follow up for UBI and PAU study arms.

90x90mm (300 x 300 DPI)

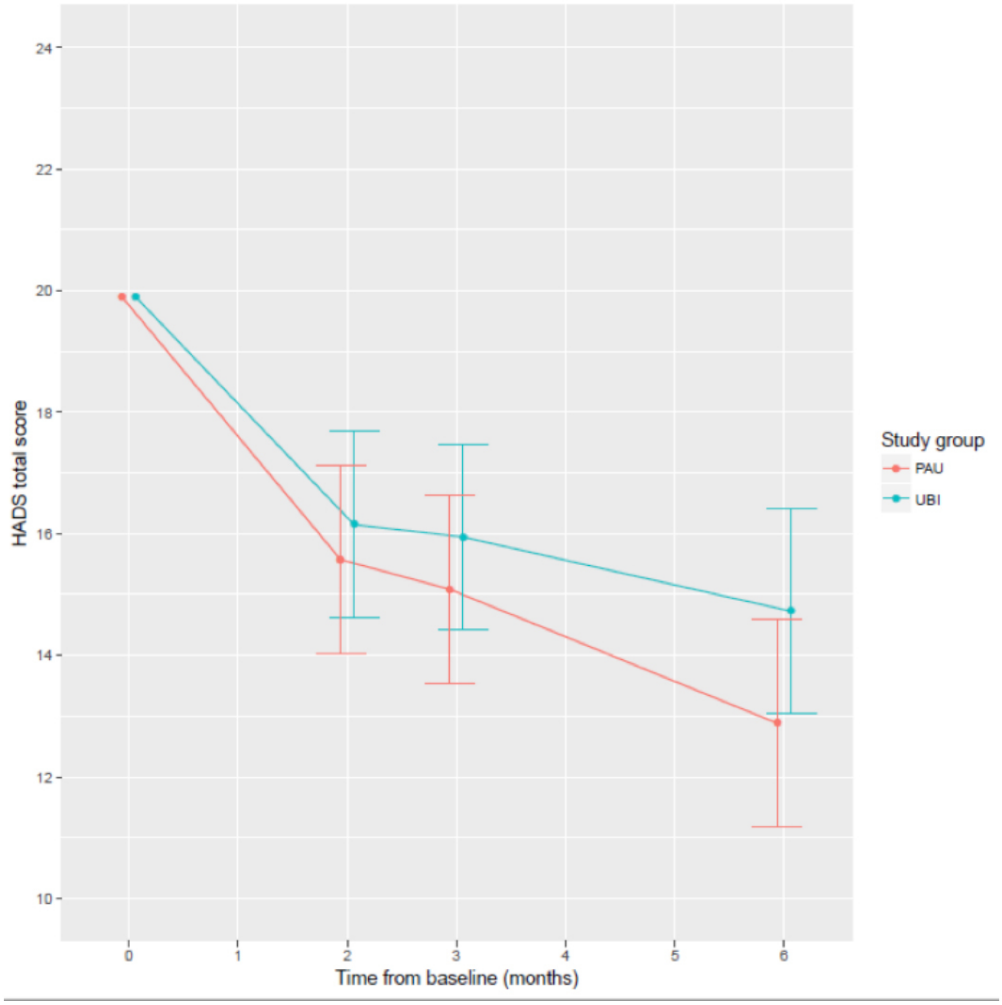
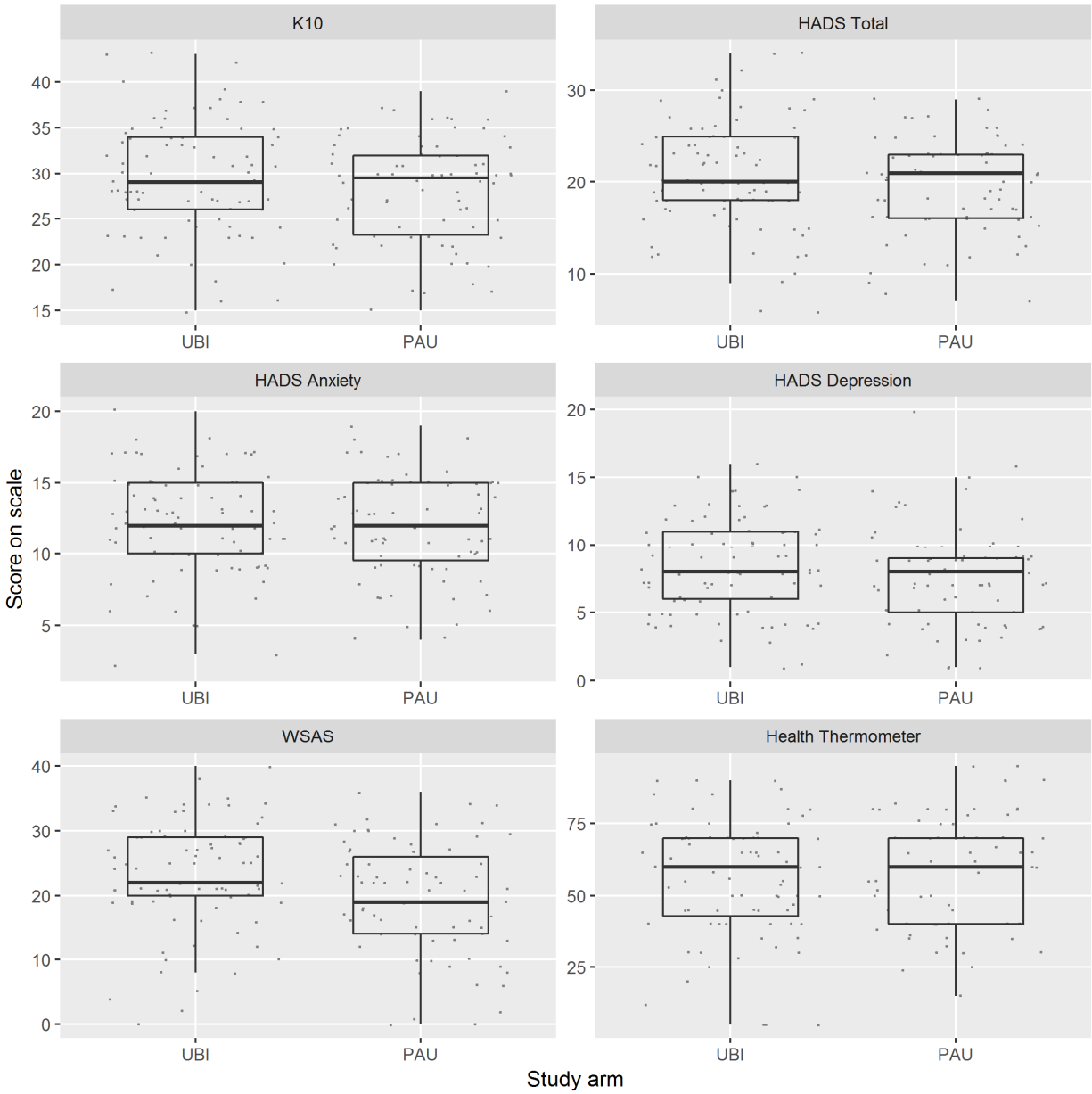


Figure 3. Mean total HADS score (95% CI) at baseline and follow up for UBI and PAU study arms.

90x90mm (300 x 300 DPI)

Supplementary Figure R1. Boxplots of baseline scores for each outcome measure (dots show each individual's score on that measure).



Supplementary Table R1. Number of patients recruited into study by GPs in UBI and PAU study arms.

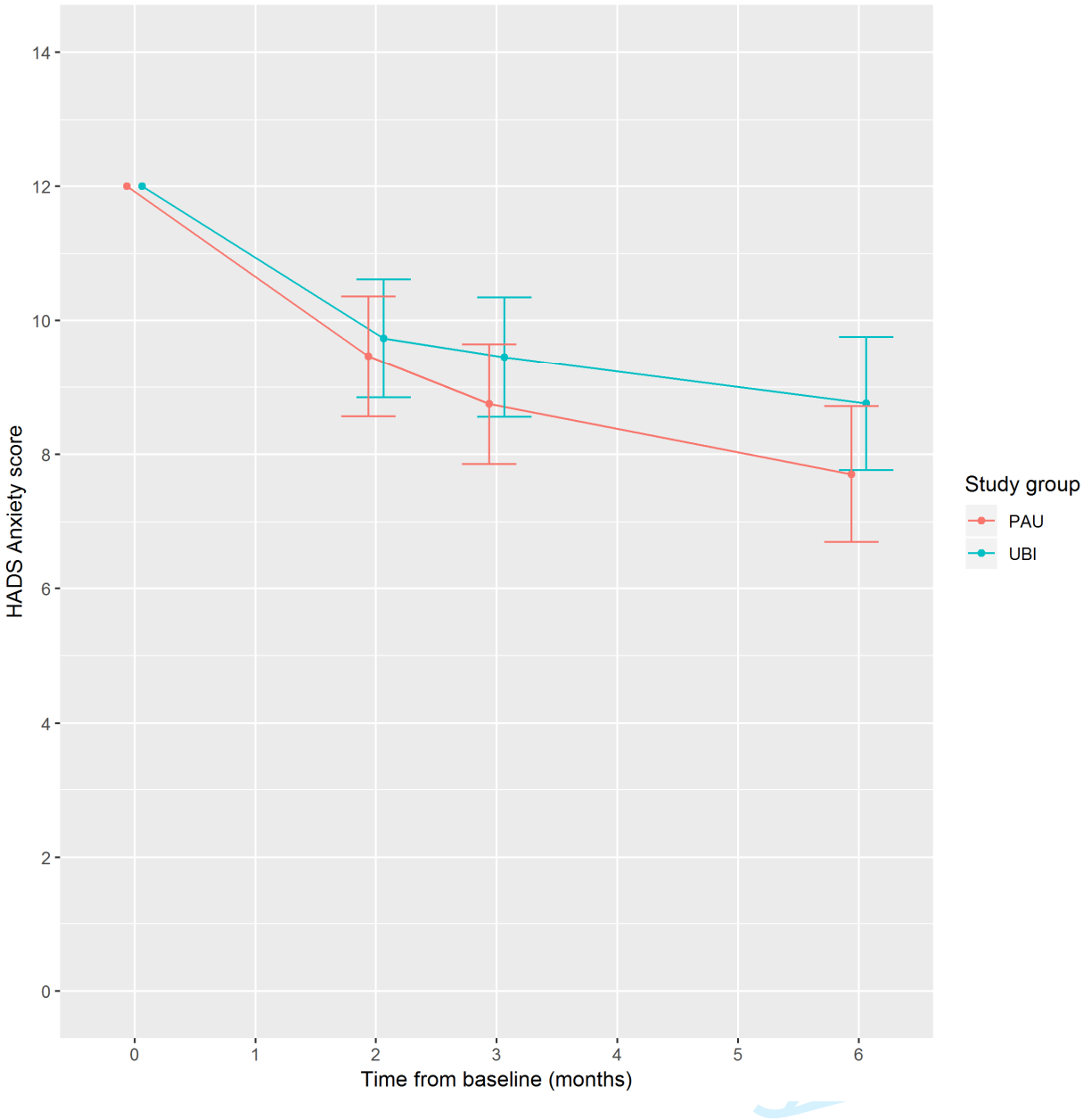
Number of patients recruited by GP	UBI (n GPs*)	PAU (n GPs*)
1	12	8
2	4	2
3	7	5
4	3	0
5	1	2
6	2	0
7	1	0
8	1	1
9	0	2
12	0	1
Total number of GPs	31	21

* Indicates the number of GPs recruiting the stated number of patients (e.g. 12 GPs in the UBI arm recruited one patient each; and five GPs in the PAU arm recruited three patients each).

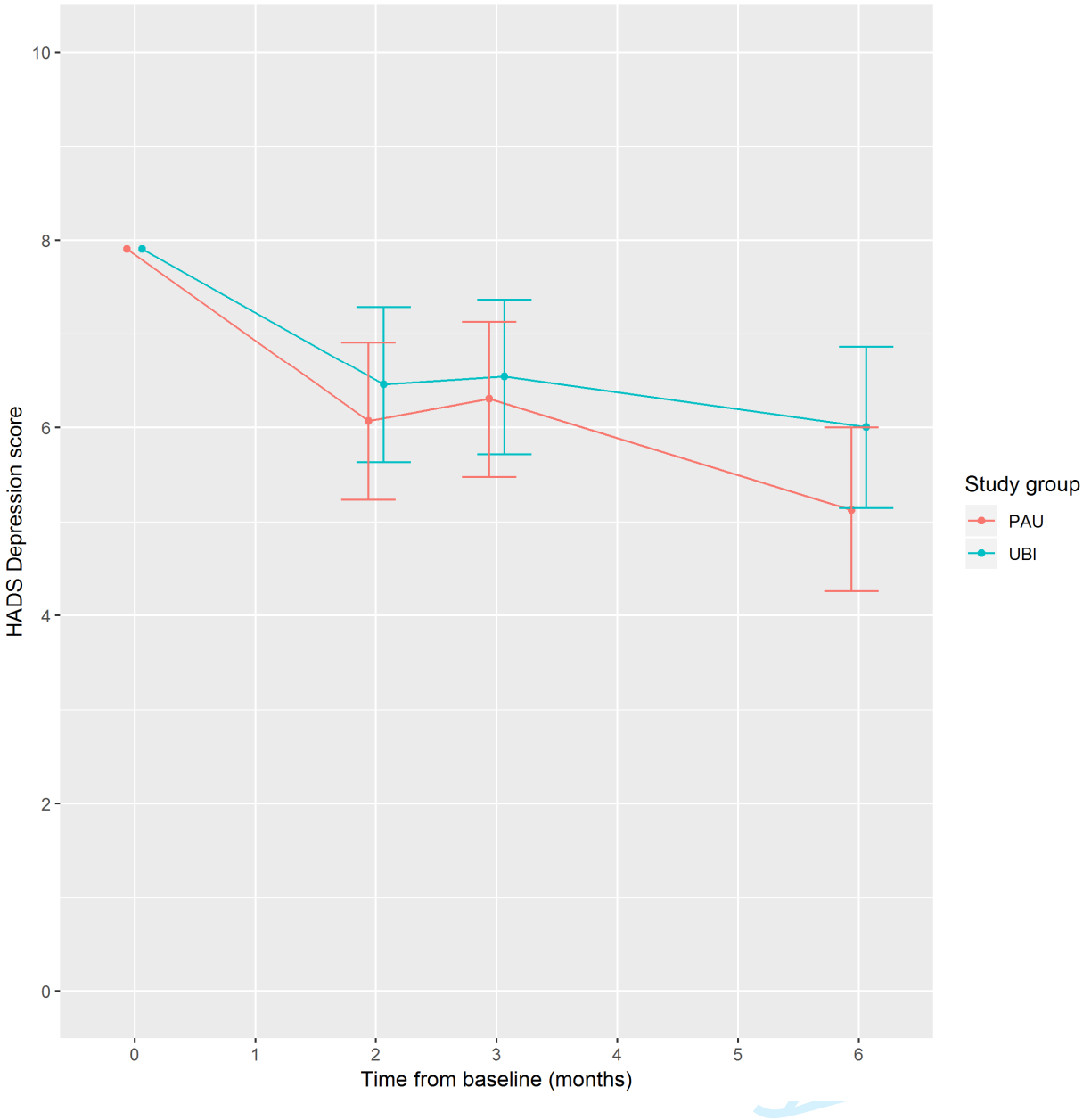
Supplementary Table R2. Mean improvements from baseline to 6 month follow-up for each outcome measure.

Outcome measure	Mean at baseline (both arms)	Mean improvement (95% CI) from baseline to 6 months	
		PAU	UBI
K10	28.8	7.6 (5.5, 9.6)	5.9 (4.0, 7.8)
HADS	19.9	7.0 (5.3, 8.7)	5.2 (3.5, 6.9)
HADS-A	12	4.3 (3.3, 5.3)	3.2 (2.2, 4.2)
HADS-D	7.9	2.8 (1.9, 3.7)	1.9 (1.0, 2.8)
WSAS	21.3	7.7 (5.7, 9.7)	7.2 (5.3, 9.2)
Health Thermometer	57.5	14.0 (9.3, 18.6)	9.0 (4.4, 13.7)

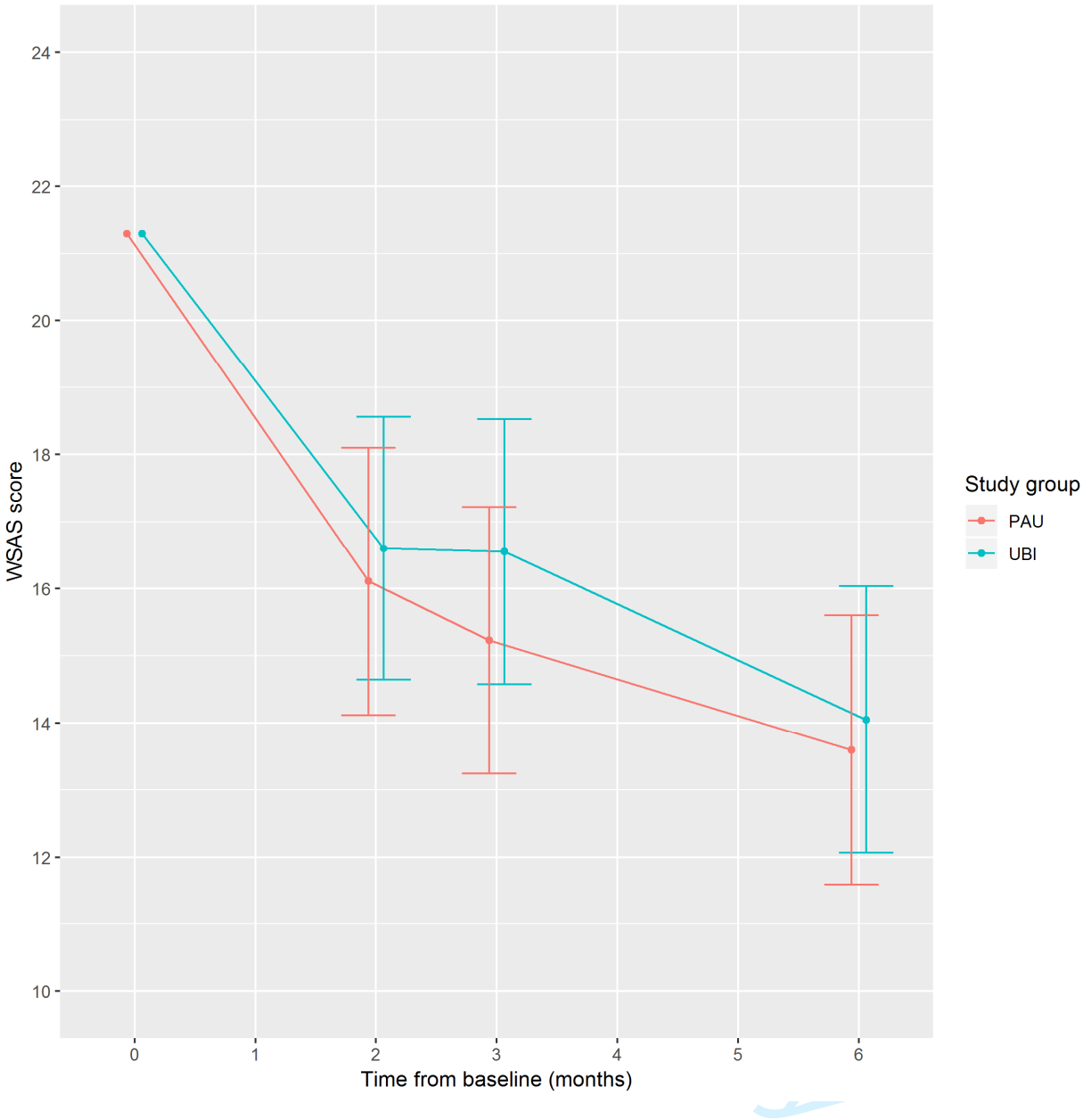
Supplementary Figure R2. Mean HADS Anxiety score (95% CI) at baseline and follow up for UBI and PAU study arms.



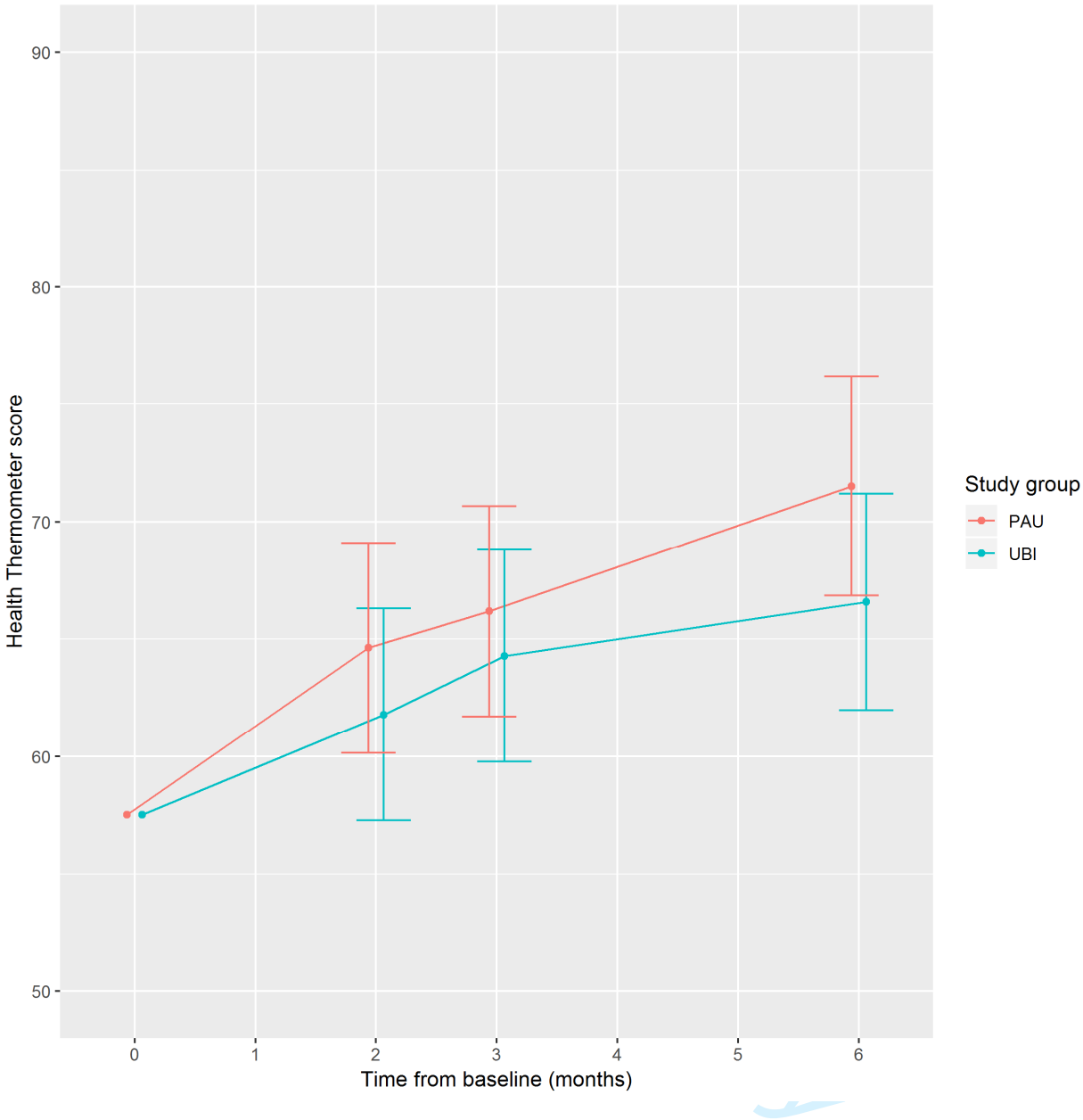
Supplementary Figure R3. Mean HADS Depression score (95% CI) at baseline and follow up for UBI and PAU study arms.



Supplementary Figure R4. Mean WSAS score (95% CI) at baseline and follow up for UBI and PAU study arms.



Supplementary Figure R5. Mean Health Thermometer score (95% CI) at baseline and follow up for UBI and PAU study arms.



Supplementary Table R3. Additional treatment received during UBI trial (from question on 6 month interview)

Type of additional treatment		UBI n (%)	PAU n (%)
Medication status during trial			
	no relevant medication	33 (51%)	34 (52%)
	on medication prior to entering trial	20 (31%)	16 (24%)
	started medication during trial	12 (18%)	16 (24%)
	did not complete question*	20	9
Extended GP consultations (n)			
	0	68 (100%)	46 (71%)
	1-2	0	8 (12%)
	3-5	0	9 (14%)
	6-10	0	2 (3%)
	did not complete question*	17	10
Counselling sessions (n)			
	0	44 (75%)	21 (36%)
	1-2	4 (7%)	13 (22%)
	3-5	2 (3%)	11 (19%)
	6-10	7 (12%)	12 (20%)
	11+	2 (3%)	2 (3%)
	did not complete question*	26	16

* Did not complete 6 month questionnaire and hence no data (UBI n=16; PAU n=9)
Did not answer Meds question at 6 months (UBI: n=4; PAU: n=1)
Did not answer Extended GP question at 6 months (UBI: n=1; PAU: n=1)
Did not answer Counselling question at 6 months (UBI: n=10; PAU: n=7)

Supplementary Methods: Calculation of intra-class correlation coefficients (ICCs) for outcome measures.

ICCs were calculated for each outcome measure in the study to summarise the impact of clustering of outcomes by GPs. These were calculated using simplified mixed linear models with random intercept terms for GPs and no adjustment for covariates. ICCs were calculated in R 3.2.3, using the lme4 package, with their 95% confidence intervals based on 1000 bootstrap resamples calculated using the bootMer() function.

ICCs were also calculated for a scenario where clustering was considered across both the individual GPs (as per the above paragraph) and the practices in which GPs worked. The difference between these two sets of estimates can be considered as the additional impact of clustering of patient responses induced by practices above and beyond clustering induced by GPs. As seen in Supplementary Table R4, there was little impact of this additional clustering on ICCs for the longer health measures (K10 and HADS: minimal difference in ICCs between the two adjustment scenarios) but there appeared to be some additional impact of practice-level clustering for the Work and Social Adjustment Scale (WSAS) and the one-item Health Thermometer.

Supplementary Table R4. Intra-class correlation coefficients (ICCs) for each outcome measure in the study.

	GP clustering only*		GP and Practice clustering**	
Outcome measure	ICC	(95% CI)	ICC	(95% CI)
K10	0.129	(0.045, 0.231)	0.139	(0.006, 0.235)
HADS (total)	0.091	(0.019, 0.189)	0.104	(<0.001, 0.185)
HADS Anxiety	0.098	(0.019, 0.198)	0.106	(<0.001, 0.190)
HADS Depression	0.140	(0.047, 0.250)	0.148	(0.018, 0.233)
WSAS	0.188	(0.081, 0.308)	0.240	(0.076, 0.348)
Health Thermometer	0.088	(0.013, 0.177)	0.135	(0.005, 0.219)

* ICC calculated using only GP-level random effects.

** ICC calculated using random effects for GPs nested within GP practices (joint clustering effect).

Reference for lmer package:

Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Supplementary Table R5. Sociodemographic and clinical characteristics at baseline by intervention arm (UBI or Practice as Usual [PAU]) and follow-up status.

Factor	Level	UBI follow-up (FU) status		PAU follow-up (FU) status	
		Lost to FU	some FU	Lost to FU	some FU
Total	All participants	15 (100%)	70 (100%)	6 (100%)	69 (100%)
Ethnicity	NZE Other	10 (67%)	51 (73%)	5 (83%)	49 (71%)
	Māori	5 (33%)	14 (20%)	1 (17%)	13 (19%)
	Pacific	0 (0%)	4 (6%)	0 (0%)	2 (3%)
	Asian	0 (0%)	1 (1%)	0 (0%)	5 (7%)
Age grp	15-24	11 (73%)	44 (63%)	6 (100%)	31 (45%)
	25-34	2 (13%)	14 (20%)	0 (0%)	15 (22%)
	35-44	1 (7%)	2 (3%)	0 (0%)	13 (19%)
	45-54	0 (0%)	5 (7%)	0 (0%)	6 (9%)
	55+	1 (7%)	5 (7%)	0 (0%)	4 (6%)
Gender	Female	7 (47%)	49 (70%)	3 (50%)	54 (78%)
	Male	8 (53%)	21 (30%)	3 (50%)	15 (22%)
NZiDep	0	3 (20%)	15 (21%)	0 (0%)	11 (16%)
	1	2 (13%)	14 (20%)	1 (17%)	16 (23%)
	2	3 (20%)	12 (17%)	2 (33%)	9 (13%)
	3	0 (0%)	10 (14%)	0 (0%)	10 (14%)
	4	2 (13%)	7 (10%)	1 (17%)	11 (16%)
	5	5 (33%)	12 (17%)	2 (33%)	12 (17%)
Education	At least secondary	15 (100%)	63 (90%)	6 (100%)	65 (94%)
	No secondary	0 (0%)	7 (10%)	0 (0%)	4 (6%)
Outcome scores at baseline		mean (sd)	mean (sd)	mean (sd)	mean (sd)
	K10	28.4 (5.9)	29.8 (6.3)	32.2 (5.3)	27.8 (5.6)
	HADS	20.2 (7.5)	20.7 (5.5)	23.0 (3.0)	19.2 (5.1)
	HADS Anxiety	11.9 (4.9)	12.2 (3.2)	13.2 (2.9)	11.7 (3.5)
	HADS Depression	8.3 (3.2)	8.5 (3.5)	9.8 (3.7)	7.5 (3.6)
	WSAS	21.7 (7.8)	23.3 (8.3)	23.8 (5.2)	19.2 (8.7)
	Health Thermometer*	57.4 (16.5)	55.0 (20.6)	50.5 (17.4)	59.5 (18.7)

* Health Thermometer: Lower scores indicate poorer health state.

Supplementary Results Text 1: Mean difference in K10 primary outcome at 6 months, adjusting only for baseline scores.

The protocol for the primary outcome (K10) analysis only specified that linear mixed model would be adjusted for baseline scores. The results from the primary analysis reported in the main paper were also adjusted for baseline sociodemographic variables (repeated in Supplementary Table R6 below from Table 3).

The analysis of K10 scores at 6 months (adjusted solely for baseline K10 scores) returned a slightly smaller mean difference between groups (poorer mean K10 score in UBI compared to PAU: difference = 1.07, 95% CI -1.67, 3.82).

This supplementary analysis draws on all participants with at least one follow-up observation. All other elements of the statistical model (accounting for clustering by GP and repeat observations for the same participant) are handled as per the main analysis (see Methods of main paper).

Supplementary Table R6. Primary outcome (K10) differences between UBI and PAU study arms at 6 months under different covariate adjustment models.

Analysis	Mean difference in K10 at 6 months (95% CI)
Analysis of all participants with some follow-up (n=139)	
Adjusted for baseline covariates *	1.68 (-1.18, 4.55)
Adjusted for baseline K10 score only**	1.07 (-1.67, 3.82)

* Result as reported in Table 3 of main paper.
** Analysis in line with specifications in protocol paper.

Supplementary Methods and Results: Sensitivity analysis to account for participants with no follow-up data.

The following analyses were implemented following initial peer-review, and were not *a priori* components of the analysis plan. Results from analyses are presented in Supplementary Table XX below, following the description of the methods and results. These sensitivity analyses aimed to consider the impact of complete loss-to-follow-up (participants no post-baseline data) on the primary outcome analysis, using two different frameworks assuming data were missing at random (MAR) or missing not at random (MNAR). A discussion of potential impacts of loss-to-follow-up on study results (attrition bias) is available in Bell et al. (2012) and discussion of missing data mechanisms can be read elsewhere (e.g. Bell et al. (2012); Newgard et al. (2015) and Sullivan et al. (2018)).

References for subsequent section:

Bell ML, Kenward MG, Fairclough DL, Horton NJ. Differential dropout and bias in randomised controlled trials: when it matters and when it may not. BMJ. 2013;346:e8668.

Newgard CD, Lewis RJ. Missing Data: How to Best Account for What Is Not Known. JAMA. 2015;314:940-1.

Sullivan TR, White IR, Salter AB, Ryan P, Lee KJ. Should multiple imputation be the method of choice for handling missing data in randomized trials? Stat Methods Med Res. 2018;27:2610-26.

van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software. 2011;45(3):1-67.

Imputation of outcomes under the Missing at Random (MAR) assumption.

Imputation was implemented using the mice package in R (van Buuren et al., 2011). All primary and secondary outcomes at all follow-up times were included in the imputation model, along with sociodemographic variables at baseline (gender/sex, age group, ethnicity, education, and NZiDep category: see Table 1 of the main paper for details about the specific sub-groups within each of these variables). Imputation was conducted separately for the intervention (UBI) and control (PAU) groups (Sullivan et al., 2018).

A total of 50 imputation datasets were created; each dataset was analysed for the primary outcome following the methods used for the main analysis in the paper (linear mixed model for K10 score at 6 months, adjusted for baseline K10 score and sociodemographic covariates). The estimates from these 50 models were then combined using Rubin's rules to produce the point estimate and 95% confidence interval (which takes into account variability in the effect estimates across all the imputed datasets.)

The intervention effect at 6 months is presented in Supplementary Table R7: under the assumption that the missing data mechanism was MAR (implemented using multiple imputation) there was a mean difference in K10 scores of 1.78 points (95% CI -0.96, 4.51; positive scores indicate better outcomes in the practice as usual [PAU] arm compared to UBI). This was almost identical to the estimates from the linear mixed model reported in Table 3 (repeated in Supplementary Table R7 for reference) which also assumed an MAR mechanism for missing data (conditional on the adjusted baseline variables in that model), but the analysis in the main results only included participants with at least one post-baseline measurement.

Imputation of outcomes under the Missing Not At Random (MNAR) assumption.

Analysis assuming that outcome values were MNAR was repeated under several conditions to explore the potential impact of different types of missing data mechanisms. These analyses all assumed that participants who did not participate in any follow-up did worse than those who participated in at least one follow-up.

In all scenarios, those who were not lost-to-follow-up (i.e. had at least one follow-up measure) kept either their original K10 scores at 6 months, or their imputed values at 6 months (for those with only partial follow-up: using the same imputed datasets as analysed under the MAR assumption). Imputation under MAR principles was considered reasonable for those with at least one follow-up measurement (but no 6-month measurement), as the follow-up measurements were all timed well after the conclusion of the core interventions delivered as part of the trial.

In MNAR Scenario 1: Individuals with no follow-up data were given a K10 score at six months set to 4 points lower than their imputed score.

In MNAR Scenario 2: Individuals with no follow-up data were given the same K10 score at six months that they had at baseline. This is effectively a "last observation carried forward" analysis for those with no follow-up data.

In MNAR Scenario 3: Individuals with no follow-up data were given a K10 score at six months that was 4 points lower than their baseline score.

The outcome analyses were again repeated on the 50 imputed datasets, and the intervention effect results combined across the resulting estimates.

While the effect sizes were slightly different from the main study result (Supplementary Table R7), these assumptions of data being MNAR had relatively minor impact on effect sizes. The most conservative result was under Scenario 1, assuming outcomes for those with no follow-up data were 4 points worse than imputed, returned a mean difference of 2.03, 95% CI -0.63, 4.70.

Note that the confidence intervals with the MNAR sensitivity analyses are likely to be conservative (i.e. not as wide as they should be) because the differences applied from the imputed or baseline values in each scenario were fixed rather than stochastic quantities (i.e. assumes that the applied difference from the imputed or baseline score was always a fixed quantity for all people).

Supplementary Table R7. Estimates of primary outcome (K10) effect size at 6 months under different assumptions of missing outcome profiles in participants with no follow-up data.

Analysis	Mean difference in K10 at 6 months (95% CI)	
Analysis of all participants with some follow-up (n=139)		
Adjusted for baseline covariates (main analysis*)	1.68	(-1.18, 4.55)
Analysis including all randomised participants (n=160)		
Imputed K10 outcome at 6 months (MAR assumption*)	1.78	(-0.96, 4.51)
Imputed K10 outcome at 6 months (MNAR assumptions*)		
1. K10 at 6m set to 4 points worse than imputed	2.03	(-0.63, 4.70)
2. K10 at 6m set to baseline score	1.45	(-0.95, 3.84)
3. K10 at 6m set to 4 point worse than baseline	1.71	(-0.95, 4.37)

* Result as reported in Table 3 of main body of paper.
* MAR (missing at random) and MNAR (missing not at random) assumptions are for the 21 participants lost to follow up (no post-baseline data).

Table 1: CONSORT 2010 checklist of information to include when reporting a cluster randomised trial

Section/Topic	Item No	Standard Checklist item	Extension for cluster designs	Page No *
Title and abstract				
	1a	Identification as a randomised trial in the title	Identification as a cluster randomised trial in the title	Title page
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)[1,2]	See table 2	In abstract
Introduction				
Background and objectives	2a	Scientific background and explanation of rationale	Rationale for using a cluster design	p. 4-5 also p. 6 (methods)
	2b	Specific objectives or hypotheses	Whether objectives pertain to the the cluster level, the individual participant level or both	p. 4
Methods				
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	Definition of cluster and description of how the design features apply to the clusters	p. 4-5
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons		n/a
Participants	4a	Eligibility criteria for participants	Eligibility criteria for clusters	p.4 (for both clusters and participants)
	4b	Settings and locations where the data were collected		p. 4
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	Whether interventions pertain to the cluster level, the individual participant level or both	p. 7
Outcomes	6a	Completely defined pre-	Whether outcome measures	p. 8

		specified primary and secondary outcome measures, including how and when they were assessed	pertain to the cluster level, the individual participant level or both	
	6b	Any changes to trial outcomes after the trial commenced, with reasons		n/a
Sample size	7a	How sample size was determined	Method of calculation, number of clusters(s) (and whether equal or unequal cluster sizes are assumed), cluster size, a coefficient of intracluster correlation (ICC or *k*), and an indication of its uncertainty	p. 8-9
	7b	When applicable, explanation of any interim analyses and stopping guidelines		n/a (no interim analysis was applied)
Randomisation:				
Sequence generation	8a	Method used to generate the random allocation sequence		p. 6
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	Details of stratification or matching if used	p. 6
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	Specification that allocation was based on clusters rather than individuals and whether allocation concealment (if any) was at the cluster level, the individual participant level or both	n/a
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	Replace by 10a, 10b and 10c	
	10a		Who generated the random allocation sequence, who enrolled clusters, and who assigned clusters to interventions	p. 6 (Recruitment and Randomisation sections)

	10b		Mechanism by which individual participants were included in clusters for the purposes of the trial (such as complete enumeration, random sampling)	p. 7 (Recruitment procedures sub-section)
	10c		From whom consent was sought (representatives of the cluster, or individual cluster members, or both), and whether consent was sought before or after randomisation	p. 6 (for GPs as the clusters) and p.7 (consent for the patients)
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how		GPs unable to be blinded (p 6) Statistician blinded during analysis (p. 9) Research assistant unable to be blinded
	11b	If relevant, description of the similarity of interventions		n/a
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	How clustering was taken into account	p.9-10 Analysis and clustering noted on p. 9
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses		p. 10
Results				
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	For each group, the numbers of clusters that were randomly assigned, received intended treatment, and were analysed for the primary outcome	Clusters (GPs) noted on p. 11, additional detail in Supplementary Table R1. Individual patients noted on p. 11, flowchart in

				Figure 1 (including who was covered in analysis)
	13b	For each group, losses and exclusions after randomisation, together with reasons	For each group, losses and exclusions for both clusters and individual cluster members	Clusters (GPs) covered on p.11 (no losses or exclusions, other than zero-recruitment which is covered in Supplementary Table R1) Patients covered in Figure 1
Recruitment	14a	Dates defining the periods of recruitment and follow-up		p. 5
	14b	Why the trial ended or was stopped		p. 5
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	Baseline characteristics for the individual and cluster levels as applicable for each group	Individual level characteristics reported in Table 1.
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	For each group, number of clusters included in each analysis	Analysis by original assigned groups (methods, p. 9) Number of participants for each analysis: Table 2, Table 3 Number of clusters (across all analyses): Supplementary Table R1
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95%	Results at the individual or cluster level as applicable and a coefficient of intracluster correlation (ICC or k) for each	Effect size and precision given in all tables and figures, and for outcomes

		confidence interval)	primary outcome	reported in body of text ICC reported on p 17 for primary outcomes, and Supplementary Table R4 for all outcomes.
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended		n/a (no binary outcomes used in study)
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory		ICCs reported on page 17 (as noted above) Information on additional treatment received presented p 17
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms[3])		n/a
Discussion				
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses		P 18-19 (recruitment not completed to planned sample size) p 20-21 (other limitations)
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	Generalisability to clusters and/or individual participants (as relevant)	p. 20-21
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence		(across discussion)
Other information				

Registration	23	Registration number and name of trial registry	p.4
Protocol	24	Where the full trial protocol can be accessed, if available	p. 4, reference list for detail
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	p. 22-23

** Note: page numbers optional depending on journal requirements*

Table 2: Extension of CONSORT for abstracts[1,2] to reports of cluster randomised trials

Item	Standard Checklist item	Extension for cluster trials
Title	Identification of study as randomised	**Identification of study as cluster randomised**
Trial design	Description of the trial design (e.g. parallel, cluster, non-inferiority)	
Methods		
Participants	Eligibility criteria for participants and the settings where the data were collected	**Eligibility criteria for clusters**
Interventions	Interventions intended for each group	
Objective	Specific objective or hypothesis	**Whether objective or hypothesis pertains to the cluster level, the individual participant level or both**
Outcome	Clearly defined primary outcome for this report	**Whether the primary outcome pertains to the cluster level, the individual participant level or both**
Randomization	How participants were allocated to interventions	**How clusters were allocated to interventions**
Blinding (masking)	Whether or not participants, care givers, and those assessing the outcomes were blinded to group assignment	
Results		
Numbers randomized	Number of participants randomized to each group	**Number of clusters randomized to each group**
Recruitment	Trial status[1]	
Numbers analysed	Number of participants analysed in each group	**Number of clusters analysed in each group**
Outcome	For the primary outcome, a result for each group and the estimated effect size and its precision	**Results at the cluster or individual participant level as applicable for each primary outcome**
Harms	Important adverse events or side effects	
Conclusions	General interpretation of the results	
Trial registration	Registration number and name of trial register	
Funding	Source of funding	

[1] Relevant to Conference Abstracts

REFERENCES

1 Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, et al. CONSORT for reporting randomised trials in journal and conference abstracts. *Lancet* 2008, 371:281-283

2 Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG at al (2008) CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Med* 5(1): e20

3 Ioannidis JP, Evans SJ, Gotzsche PC, O'Neill RT, Altman DG, Schulz K, Moher D. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med* 2004; 141(10):781-788.