Original article

Effectiveness of watchful waiting versus antidepressants for patients diagnosed of mild to moderate depression in primary care: A 12-month pragmatic clinical trial (INFAP study)

M. Iglesias-González[a,b,d], I. Aznar-Lou[c,d], M.T. Peñarrubia-María[d,e,f], M. Gil-Girbau[c,f,i], R. Fernández-Vergel[e,f], J. Alonso[d,g,h], A. Serrano-Blanco[a,c,d,*], M. Rubio-Valera[c,d,i]

[a] Parc Sanitari Sant Joan de Déu, Sant Boi de Llobregat, Spain
[b] School of Medicine, University of Barcelona, Barcelona, Spain
[c] Teaching, Research & Innovation Unit, Institut de Recerca Sant Joan de Déu, Esplugues de Llobregat, Spain
[d] Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública, CIBERESP, Madrid, Spain
[e] SAP Delta Llobregat, DAP Costa Ponent, Institut Català de la Salut (ICS), Catalonia, Spain
[f] Fundació Idiap Jordi Gol i Gurina, Barcelona, Spain
[g] Health Services Research Unit, IMIM-Hospital del Mar Medical Research Institute, Barcelona, Spain
[h] Pompeu Fabra University (UPF), Barcelona, Spain
[i] School of Pharmacy and Food Sciences, Universitat de Barcelona, Barcelona, Spain

A R T I C L E   I N F O

A B S T R A C T

Background: Although mild to moderate major depressive disorder (MDD) is one of the main reasons for consulting a general practitioner (GP), there is still no international consensus on the most appropriate therapeutic approach.
Methods: The aim of this study is to evaluate the clinical effectiveness of watchful waiting (WW) compared with the use of antidepressants (ADs) for the treatment of mild to moderate depressive symptoms in 263 primary care (PC) usual-practice patients in a 12-month pragmatic non-randomised controlled trial. Both longitudinal and per-protocol analyses were performed, through a multilevel longitudinal analysis and a sensitivity analysis.
Results: We observed a statistically significant time x treatment interaction in the severity of depression (Patient Health Questionnaire, PHQ-9) and disability (World Health Organization Disability Assessment Schedule, WHODAS) in favour of the AD group at 6 months but not at 12 months. The effect size of this difference was small. No statistically significant differences were observed between groups in severity of anxiety (Beck Anxiety Inventory, BAI) or health-related quality-of-life (EuroQol-5D, EQ-5D). Sensitivity analysis and per-protocol analysis showed no differences between the two groups in any of the evaluated scales.
Conclusions: Superiority of either treatment (WW and AD) was not demonstrated in patients treated for depression in PC after one year of follow-up.

1. Introduction

Mild and mild-moderate major depressive disorder (MDD) is one of the most important reasons for consulting a general practitioner (GP) and it will be among the three leading causes of disease burden by the year 2030 [1]. There is a lack of international consensus on the most appropriate therapeutic approach and the best strategy for implementation in Primary Care (PC). While, in some countries, pharmacological treatment with antidepressants (ADs) is recommended as an acute-phase treatment for patients with mild to moderate symptoms [2]; in most European countries a watchful waiting (WW) approach seems to be the first treatment step [3–5].

WW, also known as active monitoring or supportive care, has been described as an agreement between the clinician and the patient not to immediately treat the condition with ADs but to intermittently reassess its status during a specific follow-up period [6]. In common with most European guidelines for the treatment of depression, the Catalan Clinical Guideline [4] recommends that

* Corresponding author at: C/ Doctor Antoni Pujadas, 42, 08830, Sant Boi de Llobregat, Barcelona, Spain.
E-mail address: aserrano@pssjd.org (A. Serrano-Blanco).

patients treated in PC should be regularly visited by GPs and offered a variety of non-pharmacological interventions (low-intensity psychosocial and psychological interventions such as problem-solving techniques, active listening, counselling, brief or computerised cognitive behavioural therapy, or medical education) while closely monitoring clinical progress. Studies show that this treatment approach is limited due to barriers including the clinician's high care burden and lack of time, knowledge of psychotherapy and availability of mental health professionals for referral [7].

Our recent systematic review found only three studies properly assessing the clinical effectiveness of WW compared with ADs when mild and mild-moderate MDD was treated in PC. No statistically significant differences in effectiveness between the two treatment arms were found in any of the articles when the main analysis was conducted (longitudinal analysis) [8]. The small sample sizes may have limited the capacity of these studies to detect statistically significant differences between groups.

The aim of this study is to evaluate the clinical effectiveness of WW compared with the use of ADs for the treatment of mild-moderate depressive symptoms in real PC practice.

2. Methods

2.1. Study design

This was a 12-month follow-up naturalistic prospective controlled trial comparing patients that received AD drugs with those who received WW. The study was approved by our institution (PSSJD: EPA-24-12; IDIAP: 5013-002). The detailed study protocol has been published elsewhere [9].

2.2. Setting and participants

The study was conducted in 12 PC centres in the province of Barcelona (Catalonia, Spain) and 68 GPs participated in the recruitment of patients.

Prior to the study, GPs received a three hour-training session on the study protocol, diagnostic criteria for depression, and national guidelines for the treatment of MDD in PC.

GPs recruited patients for the study from their daily list of patients attending the practice. Eligible patients were adults ($\geq$18 years-old) diagnosed with a first or recurrent new episode (new diagnosis or relapse) of mild to moderate MDD according to the GP's clinical judgement (due to the study design, there was no need for structured clinical diagnosis through standardised assessment scales). Patients were excluded if they had taken AD medication in the previous 60 days, had taken antipsychotics, lithium or antiepileptics in the previous six months, presented psychotic or bipolar disorder, had a history of drug abuse or dependency, had cognitive impairment that prevented an assessment interview, or refused to provide signed informed consent.

2.3. Interventions

In accordance with the study's naturalistic design, GPs used their professional clinical judgment to recommend a treatment option to the patient. Patients were recommended a WW approach or pharmacological treatment with ADs.

Patients in the WW group agreed with the GP not to immediately treat the condition with ADs but to closely monitor the symptoms through a series of follow-up visits. In line with the Catalan Guideline [4], a first follow-up visit was scheduled to take place within the following 2 weeks. The guideline recommends from six to eight follow-up visits over 10–12 weeks, where GPs can consider non-pharmacological interventions. It also recommends

structured, supervised exercise programmes of moderate intensity. As part of the stepped care model, in the case that the patient's condition does not improve, the GP can intensify the treatment and initiate ADs. The number of visits following recruitment by the GP was used to monitor adherence to WW.

Patients in the AD group received pharmacological treatment with SSRIs (selective serotonin reuptake inhibitors), particularly with citalopram, sertraline, paroxetine or fluoxetine. Adherence to ADs was monitored through pharmacy records and patients' self-reported adherence (using the 4-item scale developed by Morisky et al. [10].

2.4. Outcomes

Sociodemographic data were collected on study commencement: age, gender, marital status, education and employment status.

The following outcomes were assessed at baseline, six and twelve months by an external researcher. The primary study outcome was the effectiveness of each treatment, WW or ADs, measured in terms of depression severity. This was assessed using the Patient Health Questionnaire 9-item depression module (PHQ-9) [11,12]. The scale consists of nine items scored from 0 to 3, with a summed score that ranges between 0 (no symptoms of depression) and 27 (all symptoms of depression every day): 0–4 indicates minimal symptoms; 5–9 mild depression symptoms; 10–14 moderate symptoms; 15–19 moderate-severe symptoms; and 20–24 severe symptoms.

Diagnosis according to DSM-IV criteria was not used as an inclusion criterion. However, clinical diagnosis according to DSM-IV criteria was assessed using the research version of the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) [13]. The mood and anxiety disorder modules were used.

Severity of anxiety was evaluated through the Beck Anxiety Inventory (BAI), which is a twenty-one item self-report inventory ranging from 0 (minimal level of anxiety) to 63 (severe anxiety) [14,15].

The Spanish version of the EuroQol-5D-3 L (EQ-5D) was used to measure health-related quality of life [16,17]. The EQ-5D generates a health tariff that is anchored in 1.000 (best health state) and 0.000 (being dead).

The 12-item interviewer-administered version of the World Health Organization Disability Assessment Schedule (12-item WHO-DAS 2.0) was used to assess disability [18]. Total scores range between 0 and 100, with higher scores indicating a greater degree of disability.

Cognitive representation of medication was assessed using the Beliefs about Medicine Questionnaire (BMQ) [19,20]. Total scores range between 8 and 40. Higher scores represent more negative beliefs about medicines.

Chronic physical conditions were assessed using a "yes" or "no" check-list.

See study protocol for more detailed information on the administered scales [9].

2.5. Analysis design/strategy

Both intention to treat (ITT) analysis (all patients were included in the analysis in the group to which they were allocated independently of the treatment they finally received) and per-protocol (PP) analysis (including only those patients who adhered to the treatment originally allocated) were performed. Adherence in the WW group was defined as receiving at least 3 follow-up visits and at least one of the recommended interventions (psychoeducation, problem-solving therapy and/or physical exercise). Adherence in the ADs group was defined as having a mean

medication possession ratio (number of pills filled/number of days of treatment) higher than 0.8. Sample size calculation was based on the primary study outcome: cost-effectiveness. See the study protocol [9] for detailed information.

Missing data patterns were evaluated to assess the plausibility of data missing at random [21]. To minimise bias resulting from the loss of information not following a completely random pattern, missing values were imputed using multiple imputation by chained equations. The imputation model included relevant socio-demographic and prognostic variables associated with drop-outs and outcome variables [22]. The number of imputations was calculated using a rule of thumb with respect to the fraction of missing information (M>100*FMI).

As this was a naturalistic study, GPs were allowed to recommend ADs or WW, based on their own clinical criteria, taking the patient's symptoms and preferences into account. All GPs were asked to recommend both treatments although patients receiving ADs may differ from those receiving WW in ways that could predispose them to make greater or lesser use of services and/or experience different clinical progress, leading to unbalanced comparison groups. We assessed which variables were associated with a higher probability of receiving WW. After adjustment for significant variables, the only variable that predicted the use of WW was the beliefs about medication questionnaire. Therefore, all analyses were adjusted for BMQ results.

Effect size was determined by calculating the mean difference between the two groups and then dividing the result by the pooled standard deviation.

2.6. Statistical analyses

Baseline characteristics of the groups were compared using the *t*-test for continuous variables and logistic regression for categorical variables.

Statistically significant differences between groups existed in the proportion of patients with a diagnosis of dysthymia and panic disorder with agoraphobia, as well as baseline values in the WHODAS scale and the BMQ for general medication. All analyses were adjusted for these variables, age and gender (Table 1). No other statistically significant differences existed between the two treatment arms at baseline.

Two analysis strategies were subsequently employed. Our main strategy was a multilevel longitudinal analysis with data from the same individuals grouped over three time periods (baseline, 6 months and 12 months, modelled simultaneously). To assess the effectiveness of WW, we used severity of depression, severity of anxiety, disability and quality of life as dependent variables. Time × treatment interactions were tested in all models. We considered the effect of the intervention to vary over the course of the study when the interaction was statistically significant. The model without interaction was used when the interaction term was not statistically significant. Clinical and sociodemographic characteristics of both patient and physician were tested in bivariate regression models for each of the dependent variables. In addition to the BMQ, those variables that were associated with the dependent variables in these models (p ≤ 0.05) were used to adjust longitudinal models. As pre-specified, all the analyses were adjusted for age and gender.

As in previous papers, we used a sensitivity analysis as a secondary analytical strategy [23]. We used linear regression models to assess differences between groups at 6 and 12 months. These models included severity of depression and anxiety, disability and quality of life as dependent variables and treatment as the independent variable. The adjustment strategy was similar to the main analysis as described above.

Stata MP13.1 for Windows was employed to conduct all analyses.

3. Results

3.1. Participants and drop outs

A total of 263 patients were assigned to AD (n = 145) and WW (n = 118) groups, evaluated at baseline and included in the ITT analysis. Fig. 1 shows the flowchart. Fifty-seven patients in the AD group (39%) and seventy-eight patients in the WW group (66%) did not adhere to the intervention. The rationales for non-adherence are also shown in Fig. 1. Only 88 and 40 patients in the control and intervention group, respectively, were included in the PP analysis.

3.2. Baseline characteristics

Baseline characteristics are presented in Table 1. Most participants were women (AD: 85%; WW: 78%), with mean ages of 50 and 48 years old in the AD and WW groups, respectively; and the majority of patients in both groups were married. Approximately one third of the sample had secondary education and was actively working. Thirty-one percent of the participants met DSM-IV criteria for MDD and the mean baseline severity of depression according to the PHQ-9 was 16.15 (corresponding to moderately severe symptoms). Adherence rates were 61% in the control group (AD) and 34% in the intervention group (WW). Among patients who did not receive WW as intended, 54% did not receive at least one of the recommended non-pharmacological interventions and 84% did not receive at least 3 follow-up visits. Thirty-nine percent of the patients declared having none of these two requirements.

3.3. Main outcome measures

Table 2 shows the results from the ITT longitudinal analysis at 6 and 12 months for the different scales used: severity of depression (PHQ-9), severity of anxiety (BAI), disability (WHODAS) and quality of life (EQ-5D). Our main outcome measure regarding severity of depression demonstrated a statistically significant interaction in favour of the AD group at 6 months but significance was not maintained at 12 months (p = 0.487) after treatment initiation. Effect size was very small at both 6 and 12 months (<0.04). Fig. 2 shows the development of depression severity over time for both groups.

Concerning secondary outcomes (Table 2), only a statistically significant time x group interaction in favour of ADs at 6 months in the scales testing disability (WHODAS) (p = 0.019) was observed. As with depression severity measures, this difference was not seen at 12 months (p = 0.568). Effect size was also very small at 6 months and at 12 months (<0.04). Fig. 3 shows mean values for WHODAS scores by group over time. There were no statistically significant differences at any time point in the scale scores measuring severity of anxiety or quality of life at 6 or 12 months.

There were no differences in the sensitivity analysis or the PP analysis in any of the assessment scales administered (see supplementary files).

4. Discussion

Our results show that there are no clinically relevant differences between WW and ADs in the improvement of mild-moderate depressive symptoms when diagnosed and treated in PC. There is a tendency to improve more rapidly with the use of ADs but these subtle initial differences fade at 12 months. These results match those found in previous literature [23,24], where no prominent differences in effectiveness are seen in any of the treatment arms in

Table 1
Sociodemographic characteristics of both groups at baseline.

	Antidepressants (n = 145)	Watchful Waiting (n = 118)
Gender, % women (n)	77.93% (113)	84.75% (100)
Age, mean (SD)	49.72 (14.91)	47.96 (16.02)
Civil status, % (n)		
Single	11.72 (17)	20.34 (24)
Married	62.07 (90)	59.32 (70)
Separated/divorced	19.31 (28)	14.41 (17)
Widow	6.90 (10)	5.93 (7)
Cohabitation, % (n)		
Alone	13.10 (19)	16.95 (20)
Accompanied	86.90 (126)	83.05 (98)
Education, % (n)		
No formal education	6.21 (9)	8.47 (10)
Primary studies	17.24 (25)	16.95 (20)
Secondary studies	25.52 (37)	28.81 (34)
Graduated (Bachelor's degree)	37.93 (55)	31.36 (37)
University	13.10 (19)	14.41 (17)
Working status, % (n)		
Paid employment	26.90 (39)	31.36 (37)
Unpaid work	8.97 (13)	16.10 (19)
On sick leave	18.62 (27)	15.25 (18)
Unemployed (with unemployment benefits)	19.31 (28)	15.25 (18)
Unemployed (without benefits)	13.79 (20)	11.86 (14)
Retired	10.20 (12)	10.17 (12)
Place of birth, % (n)		
Catalonia	50.34 (73)	51.72 (60)
Other Spanish regions	31.72 (46)	29.31 (34)
Other countries	17.93 (26)	18.97 (22)
Number of comorbidities, % (n)		
None	13.10 (19)	15.25 (18)
One	24.14 (35)	15.25 (18)
Two	20.69 (30)	18.64 (22)
Three	10.34 (15)	16.95 (20)
Four	13.79 (20)	14.41 (17)
Five	7.59 (11)	6.78 (8)
Six or more	10.34 (15)	12.71 (15)
Mental disorders according to DSM-IV criteria, % (n)		
Major depression	33.79 (49)	27.12 (32)
Past episode of major depression	16.55 (24)	14.41 (17)
Dysthymia	**8.28 (12)**[*]	**1.69 (2)**[*]
Panic disorder	18.62 (27)	12.71 (15)
Panic disorder with agoraphobia	**8.97 (13)**[*]	**2.54 (3)**[*]
Social phobia	4.14 (6)	3.39 (4)
Specific phobia	0.69 (1)	0 (0)
Generalized anxiety disorder	4.83 (7)	2.54 (3)
Adjustment disorder	8.97 (13)	13.56 (16)
Non affective MD	42.75 (62)	46.61 (55)
Severity of depression symptomatology (based on PHQ-9) categorized, % (n)		
Minimal	0.69 (1)	0.84 (1)
Mild	11.03 (16)	15.25 (18)
Moderate	22.07 (32)	24.58 (29)
Moderate-Severe	31.72 (46)	34.75 (41)
Severe	34.48 (50)	24.58 (29)
Baseline values of the main scales (mean score, SE)		
PHQ-9	16.68 (0.44)	15.5 (0.49)
BAI	18.99 (0.97)	17.14 (0.88)
EQ-5D	0.571 (0.017)	0.605 (0.017)
WHODAS	**39.28 (1.81)**[*]	**32.93 (1.82)**[*]
Beliefs about medication questionnaire (general medication)	**22.46 (0.44)**[*]	**24.31 (0.51)**[*]

[*] Statistically significant differences at baseline between groups (p < 0.05).

the management of mild to moderate depressive symptoms. In our recent systematic review [8], only one of the included studies reported statistically significant differences in clinical symptoms in favour of ADs in the PP analysis at 13 weeks. However, no between-group differences were observed in the same sample in the ITT analysis [24].

Despite the unexpected nature of our results and those found in the literature, some factors should be considered when interpreting the results. Treatment non-adherence, diagnostic inadequacy, and difficulties in evaluating and typifying depressive symptoms in relation to their severity are common issues in daily clinical PC

practice. These factors could limit treatment effectiveness of both WW and ADs.

Treatment non-adherence is one of the major challenges associated with the worsening of patients' clinical and economic outcomes [25,26]. Both pharmacotherapy and non-pharmacological interventions are hindered by a lack of systematic follow-up and poor adherence, and associated factors appear to be multifactorial [27]. In our study, adherence rates were low in both groups, especially in the WW group. Other pragmatic studies have also demonstrated poor adherence to treatment among outpatients with MDD, suggesting that early treatment
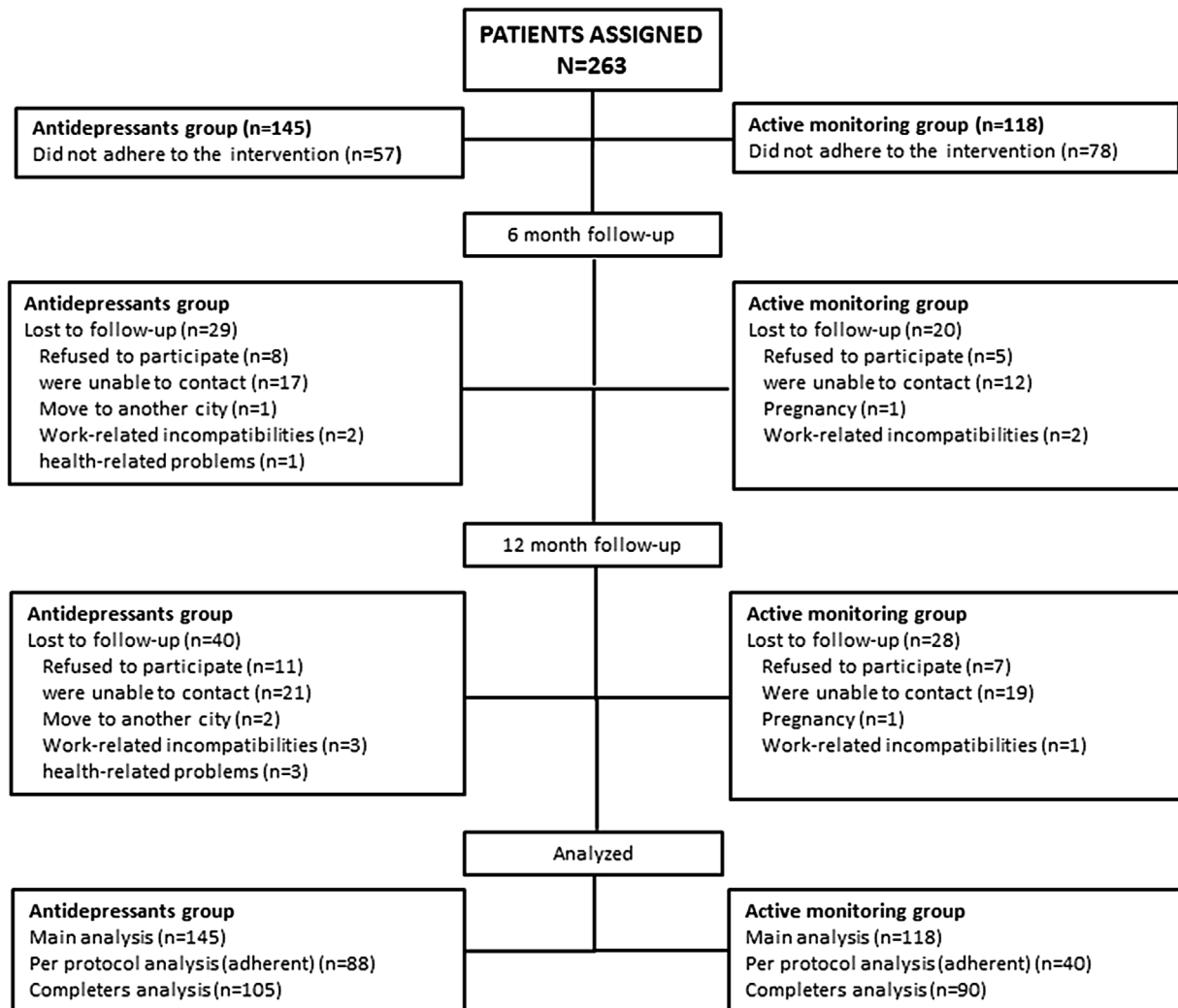
Fig. 1. Study flow chart.

optimisation and close follow-up are required to prevent long-term suffering and treatment discontinuation [28]. The naturalistic scenario of this study has strong external validity reflecting daily clinical practice and highlighting patient and GP difficulties in following clinical guidelines for the treatment of depressive disorders. Treatment non-adherence is widely recognised as one of the reasons for treatment failure in MDD and this should lead to implementation of system-level

Table 2
Multilevel model based β-coefficients (SE) and p-values for severity of depression, quality of life, disability and severity of anxiety.

	PHQ-9[A]		BAI[B]		EQ-5D[C]		WHODAS[D]	
	B-Coefficient (SE)	p-value	B-Coefficient (SE)	p-value	B-Coefficient (SE)	p-value	B-Coefficient (SE)	p-value
Group (WW)	−0.362 (0.745)	0.612	−1.161 (0.922)	0.208	−0.017 (0.017)	0.305	−2.025 (1.843)	0.272
Time								
6 months	−8.207 (0640)	≤0.001	−7.404 (0.725)	≤0.001	0.088 (0014)	≤0.001	−18.054 (1.766)	≤0.001
12 months	−7.685 (0682)	≤0.001	−5.948 (0.751)	≤0.001	0.101 (0015)	≤0.001	−16.354 (1.835)	≤0.001
TimexGroup								
6 months	1.861 (0.947)	**0.049**	Ni	Ni	Ni	Ni	6.155 (2.625)	**0.019**
12 months	0.687 (0.988)	0.487	Ni	Ni	Ni	Ni	1.528 (2.673)	0.568
_cons	17.710 (1.486)	≤0.001	23.158 (3.288)	≤0.001	0.510 (0053)	≤0.001	41.348 (4.267)	≤0.001

MDD: Major Depressive Disorder, SE: Standard Error, Ni: No interaction, WW: Watchful Waiting.
The bold values statistically significant differences between groups (p < 0.05).
[A] Adjusted for sex, age, civil status, educational qualifications, working status, number of comorbidities, baseline WHODAS mean value, dysthymia, adjustment disorder, panic disorder, panic disorder with agoraphobia, social phobia and beliefs about medication.
[B] Adjusted for sex, age, civil status, cohabitation, educational qualifications, working status, number of comorbidities, cardiovascular comorbidities, baseline WHODAS mean value, previous MDD, adjustment disorder, panic disorder, social phobia, dysthymia, panic disorder with agoraphobia and beliefs about medication.
[C] Adjusted for sex, age, educational qualifications, working status, number of comorbidities, respiratory comorbidities, digestive comorbidities, baseline WHODAS mean value, previous MDD, dysthymia, anxiety disorder with agoraphobia and beliefs about medication.
[D] Adjusted for time, sex, age, educational qualifications, working status, number of comorbidities, digestive comorbidities, baseline WHODAS mean value, dysthymia, anxiety disorder with agoraphobia and beliefs about medication.
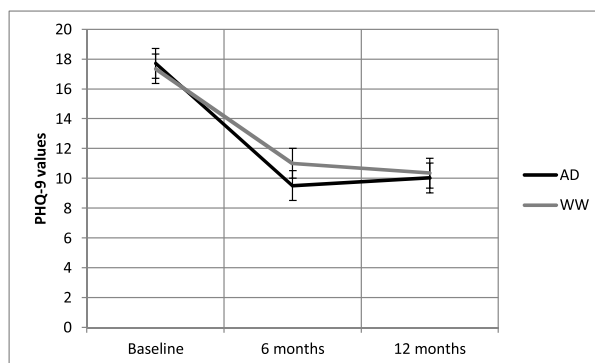
	Baseline	6 months	12 months
AD (mean score (SE))	17.71 (1.49)	9.50 (1.52)	10.03 (1.54)
WW (mean score (SE))	17.35 (1.56)	11.00 (1.58)	10.35 (1.60)

AD = Antidepressant group

WW = Watchful waiting group

Fig. 2. Development of severity of depressive symptoms (Mean PHQ9 score and SE) by group based on the regression model.
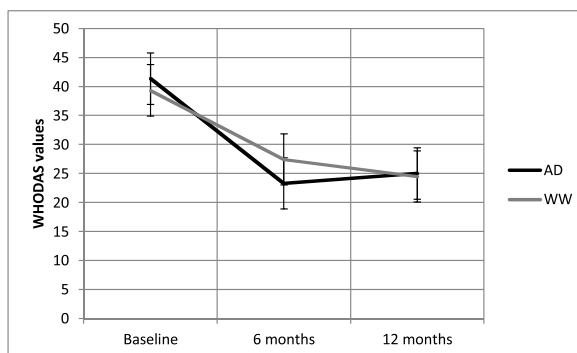
	Baseline	6 months	12 months
AD (mean score (SE))	41.35 (4.27)	23.29 (4.33)	24.99 (4.38)
WW (mean score (SE))	39.32 (4.37)	27.42 (4.43)	24.50 (4.40)

**AD = Antidepressant group

**WW = Watchful waiting group

Fig. 3. Development of disability symptoms (Mean values for WHODAS score and SE) by group based on the regression model.

interventions to improve treatment response and perceived health benefits [29].

Qualitative research has explored the conversational influences on physician decision-making about treatment for depression. The results of the study point out that patients' preferences and conceptual models of depression treatment play an important role in physician decision making. Though patient cues regarding their beliefs and preferences were presented in a subtle form, physicians appeared remarkably sensitive to these cues [30]. Recent literature [31] also suggests that exploring patient knowledge and views on depression and ADs seems to be vital when approaching MDD treatment. Development of patient-centred care systems, involving the patient more actively in treatment decision-making could improve adherence and, consequently, treatment efficacy [32].

Other studies suggest implementing psychosocial and multi-disciplinary interventions to improve early adherence to treatment [28,33]. This first step is part of the collaborative chronic care model which the literature consistently supports as a type of multi-component intervention to improve physical and mental outcomes for individuals with mental disorders. The collaborative

care model includes a range of interventions of varying intensity, ranging from simple telephone interventions to more complex interventions involving structured psychosocial interventions. Collaborative care has proved to be more effective than standard care in improving depression outcomes in the short and longer terms [34,35], even by successfully treating depressive symptoms among older adults presenting with subthreshold depression who do not meet clinical criteria for MDD [36]. These multifaceted interventions infer a greater role for nonmedical specialists such as nurse practitioners working in conjunction with PC physicians and mental health specialists. The figure of a care manager for maintaining continuous contact with patients with depression is part of this model and has positive effects on depression course, return to work, remission frequency, antidepressant frequency, and quality of life compared to usual care [37]. In Spain, there have been few collaborative care programs which have also demonstrated better clinical outcomes in patients with MDD in PC settings [38]. However, the lack of resources and training are still making it difficult to implement this complex model in real clinical PC practice.

MDD diagnostic adequacy should also be taken into consideration. In our sample, only 27.12% of patients in the WW group and 33.78% of the patients in the AD group met DSM-IV criteria for a major depressive episode. The remaining patients satisfied criteria for other affective disorders or did not meet criteria for any affective or anxiety disorder in the SCID-I. A large meta-analysis published in 2009 by Mitchell et al. [39] with 50,371 patients diagnosed with depression by GPs revealed adequate diagnosis in 47.3% of cases, with more false positives than either missed or identified cases. Other studies performed in our country show even more negative results. In the DASMAP study, only a quarter of cases diagnosed as depressed by GPs were confirmed by a structured diagnosis [40]. Mitchell et al. highlighted the importance of re-assessment as a way to improve diagnostic accuracy. However, it is not clear that a correct MDD diagnosis would improve patient response to treatment. Previous research has found that patients suffering from other depressive disorders (minor depression, depressive disorder not otherwise specified) may not be qualitatively different as many subjects may later develop a MDD [41–43]. Moreover, it seems that the degree of benefit of AD medication compared with placebo in patients diagnosed as depressed following diagnostic criteria increases with severity of depression symptoms and is minimal or non-existent in patients with mild or moderate symptoms [44]. In the subgroup of patients with mild-moderate depressive symptoms, diagnostic inadequacy might not underlie lack of response to treatment. However, it is still indicative of the lack of time and resources for diagnosing and reassessing depressed patients in a PC setting.

Depression severity has been directly associated with valid diagnosis of MDD in many studies [45,46]. A meta-analysis showed that only one in three people are correctly diagnosed when presenting with distress and mild depression in PC [47]. Patients in our sample had a mean severity PHQ-9 score corresponding to moderately-severe symptom intensity. Patients with less severe depressive symptoms (and lower PHQ-9 scores) may not contact PC services and could remain undiagnosed as they did not consult a GP. Active efforts to identify these patients might be necessary to prevent future development of more severe forms of depression [41].

Patients with somatic comorbidities may also need different strategies when diagnosed and treated for depression [46]. Most patients in our sample had at least 2 somatic comorbidities (63% of patients in the control group and 69% of patients in the intervention group) that might have affected in GPs' evaluation of depressive symptoms and the way they approach them in daily clinical practice.

In this context, we do understand the discrepancies among different clinical guidelines that adapt their recommendations in response to existing resources (Departament de Salut Generalitat de Catalunya, 2010; NICE, 2009; APA, 2010). In the UK, for instance, there is a growing number of experienced counsellors developing specific psychological and psychosocial interventions in PC practice [3]. Although clinical guidelines also recommend the same non-pharmacological approach in Spain, there is an acknowledged lack of resources and insufficient coordination between PC and Mental Health Services [48,49]. In this context, we contemplate the need to adjust clinical guidelines to real clinical practice and to devote more time and resources to improving health education, patient-centred care programmes, adherence to treatment and communication with mental health specialists. The effectiveness of each treatment strategy could be formally evaluated to provide more reliable information on the genuine, direct effectiveness of treatments.

4.1. Strengths and limitations

The naturalistic design of the study is considered a meaningful strength because it gives us a very clear view of how mild to moderate depression is approached in PC, detecting where weaknesses in the system are and allowing us to suggest where improvements could be made. However, it could also be considered as a limitation because a non-randomisation method was used that may have generated unbalanced groups. To minimise this, we adjusted our analysis with the variables that were associated with higher probability of receiving WW and the variables that differed between groups at baseline.

Another factor related to a naturalistic design concerns the low concordance rate between the clinical diagnosis made by the GP and the diagnosis according to SCID-I criteria, which may prevent us from evaluating the effectiveness of both treatments on pure MDD, but on mild to moderate depressive symptoms which do not meet MDD criteria. In our sample, the majority of patients present with subthreshold depressive symptoms or mild to moderate depression. Subsequently, generalization to MDD is not possible due to inadequate diagnosis.

Reasons for refusal or non-participation of the patients invited to take part of the trial were not recorded, not allowing us to determine how typical these participants are.

Adherence to protocol was low, which may have influenced the results obtained. To minimise this, we performed sensitivity and PP analyses.

GP participation was voluntary and therefore, GP participants may have been more motivated in mental health than their colleagues. As a consequence, their daily practice might be better adapted to the needs of these patients than standard PC practice overall.

5. Conclusion

Based on the results obtained in our study, which are in line with the existing literature, there is not enough evidence to support the superiority of WW or ADs in approaching mild to moderate depressive symptoms in PC. More research is still needed to inform GPs or policy-makers before either approach is strongly indicated. This research should also aim to determine which elements, such as an increase in GP resources (education, time, supporting counsellors and health advisors), diagnostic adequacy and/or treatment adherence, could improve the treatment of depression in PC. Practical implications would include patient decision-making strategies, GP training on how to explore the patients and other interview techniques, improve active detection of mild depression and the management of non-pharmacological treatment approaches.

Conflicts of interest

Acknowlegdment

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.eurpsy.2018.06.005.

References

[1] Paykel E.S., Brugha T, Fryers T. Size and burden of depressive disorders in Europe. Eur Neuropsychopharmacol 2005;15(August (4)):411–23.

[2] American Psychiatry Association (APA). Practice guideline for the treatment of patients with major depressive disorder. APA; 2010. . (Accessed November 2017) https://psychiatryonline.org/pb/assets/raw/sitewide/practice_guidelines/guidelines/mdd.pdf.

[3] National institute for health and clinical excellence (NICE). Depression in adults: recognition and management clinical guideline. NICE; 2009. . (Accessed November 2017) nice.org.uk/guidance/cg90.

[4] Adaptació al model sanitari català de la guia de pràctica clínica sobre el maneig de la depressió major en l'adult. Barcelona: Agència d'Informació, Avaluació i Qualitat en Salut, Pla director de Salut mental i addiccions. Departament de Salut, Generalitat de Catalunya; 2010. . (Accessed November 2016) http://aquas.gencat.cat/web/.content/minisite/aquas/publicacions/2010/pdf/adaptaciogpc_depressio_aiaqs_2010ca.pdf.

[5] van Avendonk MJP, van Weel-Baumgarten EM, van der Weele GM, Wiersma T, Burgers JS. Summary of the Dutch College of General practitioners' practice guideline "Depression". Ned Tijdschr Geneeskd 2012;156(38)A5101 [Internet].

[6] Hegel MT, Oxman TE, Hull JG, Swain K, Swick H. Watchful waiting for minor depression in primary care: remission rates and predictors of improvement. Gen Hosp Psychiatry 2006;28(May (3)):205–12.

[7] Meredith LS, Cheng WJY, Hickey SC, Dwight-Johnson M. Factors associated with primary care clinicians' choice of a watchful waiting approach to managing depression. Psychiatr Serv 2007;58(1):72–8.

[8] Iglesias-González M, Aznar-Lou I, Gil-Girbau M, Moreno-Peral P, Peñarrubia-María MT, Rubio-Valera M, et al. Comparing watchful waiting with antidepressants for the management of subclinical depression symptoms to mild–moderate depression in primary care: a systematic review. Fam Pract 2017;19(June (19)):155–62.

[9] Rubio-Valera M, Beneitez I, Peñarrubia-María MT, Luciano JV, Mendive JM, McCrone P, et al. Cost-effectiveness of active monitoring versus antidepressants for major depression in primary health care: a 12-month non-randomized controlled trial (INFAP study). BMC Psychiatry 2015;15:63.

[10] Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. Med Care 1986;24(January (1)):67–74.

[11] Diez-Quevedo C, Rangil T, Sanchez-Planell L, Kroenke K, Spitzer RL. Validation and utility of the patient health questionnaire in diagnosing mental disorders in 1003 general hospital Spanish inpatients. Psychosom Med 2001;63(4):679–86.

[12] Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary care evaluation mental disorders. Patient health questionnaire. JAMA 1999;282(November (18)):1737–44.

[13] First Michael B, Spitzer Robert L, Miriam Gibbon, Williams Janet BW. Structured clinical interview for DSM-IV-TR axis i disorders. Research version, non-patient edition. (SCID-I/NP) New York: biometrics research. New York State Psychiatric Institute; 2002 November.

[14] Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. J Consult Clin Psychol 1988;56(December (6)):893–7.

[15] Sanz J, García-Vera MP, Espinosa R, Fortún M, Vázquez C. Adaptación española del Inventario para la Depresión de Beck-II (BDI-II): 3. Propiedades psicométricas en pacientes con trastornos psicológicos. Clínica y Salud 2005;16(2):121–42.

[16] EuroQol Group. EuroQol–a new facility for the measurement of health-related quality of life. Health Policy (Amsterdam Netherl) 1990;16(3):199–208.

[17] Badia X, Schiaffino A, Alonso J, Herdman M. Using the EuroQoI 5-D in the Catalan general population: feasibility and construct validity. Qual Life Res 1998;7(May (4)):311–22.

[18] Luciano JV, Ayuso-Mateos JL, Fernandez A, Aguado J, Serrano-Blanco A, Roca M, et al. Utility of the twelve-item World Health Organization disability assessment schedule II (WHO-DAS II) for discriminating depression "caseness" and severity in Spanish primary care patients. Qual Life Res 2010;19(February (1)):97–101.

[19] Beléndez-Vázquez M, Horne R. Evaluación de las creencias sobre el tratamiento: validez y fiabilidad de la versión española del beliefs about medicines questionnaire 1. Int J Clin Health Psychol 1697;7(3):767–79.

[20] Horne R, Weinman J, Hankins M. The beliefs about medicines questionnaire: the development and evaluation of a new method for assessing the cognitive representation of medication. Psychol Health 1999;14(January (1)):1–24.

[21] Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009;29(June (338))b2393.

[22] White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Stat Med 2011;30(February (4)):377–99.

[23] Kendrick T, Chatwin J, Dowrick C, Tylee A, Morriss R, Peveler R, et al. Randomised controlled trial to determine the clinical effectiveness and cost-effectiveness of selective serotonin reuptake inhibitors plus supportive care, versus supportive care alone, for mild to moderate depression with somatic symptoms in primary care. Health Technol Assess 2009;13(22)1–159 iii–iv, ix–xi.

[24] Hermens ML, van Hout HP, Terluin B, Adèr HJ, Penninx BW, van Marwijk HW, et al. Clinical effectiveness of usual care with or without antidepressant medication for primary care patients with minor or mild-major depression: a randomized equivalence trial. BMC Med 2007;5(December (1)):36.

[25] López-Torres J, Párraga I, Del Campo JM, Villena A, ADSCAMFYC Group. Follow up of patients who start treatment with antidepressants: treatment satisfaction, treatment compliance, efficacy and safety. BMC Psychiatry 2013;13(December (1)):65.

[26] Ho SC, Chong HY, Chaiyakunapruk N, Tangiisuran B, Jacob SA. Clinical and economic impact of non-adherence to antidepressants in major depressive disorder: a systematic review. J Affect Disord 2016;15(March (193)):1–10.

[27] van Servellen G, Heise BA, Ellis R. Factors associated with antidepressant medication adherence and adherence-enhancement programmes: a systematic literature review. Ment Health Fam Med 2011;8(December (4)):255–71.

[28] Sirey JA, Banerjee S, Marino P, Bruce ML, Halkett A, Turnwald M, et al. Adherence to depression treatment in primary care. JAMA Psychiatry 2017;74 (November (11)):1129.

[29] Simon GE, Johnson E, Stewart C, Rossom RC, Beck A, Coleman KJ, et al. Does patient adherence to antidepressant medication actually vary between physicians? J Clin Psychiatry 2017;(October):24.

[30] Karasz A, Dowrick C, Byng R, Buszewicz M, Ferri L, Hartman TCO, et al. What we talk about when we talk about depression: doctor-patient conversations and treatment decision outcomes. Br J Gen Pract 2012;62(January (594)):55–63.

[31] Jaffray M, Cardy AH, Reid IC, Cameron IM. Why do patients discontinue antidepressant therapy early? A qualitative study. Eur J Gen Pract 2014;20 (September (3)):167–73.

[32] van Schaik DJF, van Marwijk HWJ, van der Windt AWM, Beekman ATF, de Haan M, van Dyck R. De effectiviteit van psychotherapie in de eerste lijn bij patiënten met een depressieve stoornis. [Effectiveness of psychotherapy for depressive disorder in primary care. A systematic review.]. Tijdschr Psychiatr 2002;44 (9):609–19.

[33] Rubio-Valera M, Serrano-Blanco A, Magdalena-Belío J, Fernández A, García-Campayo J, March Pujol M, et al. Effectiveness of pharmacist care in the improvement of adherence to antidepressants: a systematic review and meta-analysis. Ann Pharmacother 2011;45(January (1)):39–48.

[34] Gilbody S, Bower P, Fletcher J, Richards D, Sutton AJ. Collaborative care for depression. Arch Intern Med 2006;166(November (21)):2314.

[35] Woltmann E, Grogan-Kaylor A, Perron B, Georges H, Kilbourne AM, Bauer MS. Comparative effectiveness of collaborative chronic care models for mental health conditions across primary, specialty, and behavioral health care settings: systematic review and meta-analysis. Am J Psychiatry 2012;169 (August (8)):790–804.

[36] Gilbody S, Lewis H, Adamson J, Atherton K, Bailey D, Birtwistle J, et al. Effect of collaborative care vs usual care on depressive symptoms in older adults with subthreshold depression. JAMA 2017;317(February (7)):728.

[37] Björkelund C, Svenningsson I, Hange D, Udo C, Petersson E-L, Ariai N, et al. Clinical effectiveness of care managers in collaborative care for patients with depression in Swedish primary health care: a pragmatic cluster randomized controlled trial. BMC Fam Pract 2018;19(December (1)):28.

[38] Aragonès E, Lluís Piñol J, Caballero A, López-Cortacans G, Casaus P, Maria Hernández J, et al. Effectiveness of a multi-component programme for managing depression in primary care: a cluster randomized trial. The INDI project. J Affect Disord 2012;142(December (1–3)):297–305.

[39] Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. Lancet 2009;374:609–19.

[40] Fernández A, Pinto-Meza A, Bellón JA, Roura-Poch P, Haro JM, Autonell J, et al. Is major depression adequately diagnosed and treated by general practitioners? Results from an epidemiological study. Gen Hosp Psychiatry 2010;32(2):201–9.

[41] Cuijpers P, Smit F. Subthreshold depression as a risk indicator for major depressive disorder: a systematic review of prospective studies. Acta Psychiatr Scand 2004;109(May (5)):325–31.

[42] Fogel J, Eaton WW, Ford DE. Minor depression as a predictor of the first onset of major depressive disorder over a 15-year follow-up. Acta Psychiatr Scand 2006;113(January (1)):36–43.

[43] Serrano-Blanco A, Pinto-Meza A, Suárez D, Peñarrubia MT, Haro JM, ETAPS Group. (Efectivitat del Tractament antidepressiu en Atenció Primària de Salut). Cost-utility of selective serotonin reuptake inhibitors for depression in primary care in Catalonia. Acta Psychiatr Scand 2006;114(December (s432)):39–47.

[44] Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, Shelton RC, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. JAMA 2010;303(January (1)):47–53.

[45] Castro-Rodríguez JI, Olariu E, Garnier-Lacueva C, Martín-López LM, Pérez-Solà V, Alonso J, et al. Diagnostic accuracy and adequacy of treatment of depressive and anxiety disorders: a comparison of primary care and specialized care patients. J Affect Disord 2015;172(February (1)):462–71.

[46] Nuyen J, Volkers AC, Verhaak PFM, Schellevis FG, Groenewegen PP, Van den Bos GAM. Accuracy of diagnosing depression in primary care: the impact of chronic somatic and psychiatric co-morbidity. Psychol Med 2005;35(August (8)):1185–95.

[47] Mitchell AJ, Rao S, Vaze A. Can general practitioners identify people with distress and mild depression? A meta-analysis of clinical accuracy. J Affect Disord 2011;130(April (1–2)):26–36.

[48] Calderón C, Balagué L, Iruin Á, Retolaza A, Belaunzaran J, Basterrechea J, et al. Colaboración atención primaria-salud mental en la asistencia a pacientes con depresión: evaluación de una experiencia piloto. Atención Primaria 2016;48 (June (6))356–65 [Internet].

[49] Latorre Postigo JM, López-Torres Hidalgo J, Montañés Rodríguez J, Parra Delgado M. Percepción de la demanda y necesidades de formación en salud mental de los médicos de atención primaria. Atención Primaria 2005;36(June (2)):85–92.