
Article:
Firth, N., Barkham, M. and Kellett, S. (2015) The clinical effectiveness of stepped care systems for depression in working age adults: A systematic review. Journal of Affective Disorders, 170. 119 - 130. ISSN 0165-0327

https://doi.org/10.1016/j.jad.2014.08.030

The Clinical Effectiveness of Stepped Care Systems for Depression in Working Age Adults: A Systematic Review

Nick Firth[a] (corresponding author)[1]

Michael Barkham[b]

&

Stephen Kellett[b]

[a]Clinical Psychology Unit, University of Sheffield, Western Bank, Sheffield S10 2TN, UK

[b]Centre for Psychological Services Research, University of Sheffield, Western Bank, Sheffield S10 2TN, UK

A version of this article is published in Journal of Affective Disorders, 170 (2015) 119–130, http://dx.doi.org/10.1016/j.jad.2014.08.030

[1]nick.firth@gmail.com

Abstract

Background: Stepped care service delivery models involve treatments that become increasingly intense through successive steps, with patients re-assigned via pre-defined decision criteria. This article reviews the clinical effectiveness of stepped care systems for depression in working age adults.

Methods: Systematic literature review of quantitative clinical outcome evidence comprising 14 controlled and uncontrolled studies meeting specified criteria. Principal outcomes were (a) recovery rates, defined as patients no longer meeting clinical cut-off criteria for the specific outcome measure and (b) treatment response rates, defined as a 50% decrease in outcome measure score.

Results: Stepped care systems had recovery rates ranging predominantly between 40-60% and response rates approximating 60%. Studies comparing stepped care with usual/enhanced usual care tended to find significant differences favouring stepped care. The median recovery odds ratio was 1.31 (interquartile intervals of 1.05 and 1.66; k = 7 studies). The median comparative Cohen's d effect size estimate was 0.41 (interquartile intervals 0.25 and 0.45; k = 5 studies).

Limitations: The inclusion of uncontrolled studies could be seen as reducing the overall quality of evidence and a meta-analysis was not included due to limitations with the available data.

Conclusions: Evidence suggested that stepped care interventions for depression are at least as effective as usual care. However, the clinical and organisational superiority of stepped care is yet to be scientifically verified. Differential benefits of stepped care may ultimately depend on service quality. Further research investigating and comparing the specific components and configurations of stepped care interventions is indicated.

Keywords: Depression, Review, "Stepped Care", "Stepped-Care", Intervention, Effectiveness

Introduction

Background

Depression is a common mental health problem with prevalence estimates of around 10% (12 months) and 15% (lifetime) (Kessler et al., 2003; Singleton et al., 2001). Depression creates a high degree of associated burden/disability (World Health Organisation, 2001), reduced quality of life and increased functional impairment (Wells et al., 1989; Von Korff et al., 1993). The economic impact of depression is significant due to inability to work (Centre for Economic Performance, 2006; Cuijpers et al., 2007; Layard, 2006). Despite the emotional suffering created by depression, adherence to evidence-based treatment protocols has been shown to be poor (Rollman et al., 2006; Wang et al., 2007). A range of service delivery models have been developed to try to meet the needs of depressed patients comprising disease management (Hunter & Fairfield, 1997), collaborative care (Katon et al., 1997; Simon, 2006) and stepped care (Davison, 2000; Haaga, 2000; Sobell & Sobell, 2000). Stepped care service delivery models are defined by differing treatment components being available at different specified levels of intensity or 'steps'. Two key principles underpin stepped care: (1) patients initiate treatment at the least restrictive (or least intensive) step shown to be effective for their problem, and (2) the system is self-correcting in that progress is monitored and patients stepped up/down depending on specific criteria (usually lack of clinical response or increasing need/risk).

Rationale and Objective

Stepped care service design is advocated in the National Institute for Health and Clinical Excellence (NICE) guidelines for depression, anxiety and obsessive compulsive disorder (NICE, 2005, 2009, 2011). Stepped care has begun to be embraced for eating and drug/alcohol disorders (Jaehne et al., 2012; Kay-Lambkin et al., 2010; Wilson et al., 2000) and evidence for stepped care in older adult and child populations is increasing (van der Leeden et al., 2011; van't Veer-Tazelaar

et al., 2009). The challenge of providing evidence-based treatments for mental health difficulties in developing countries (Patel, 2007; Siddiqi & Siddiqi, 2007) has resulted in stepped care being advocated as the organisational model of choice (Chatterjee et al., 2008). Improving access to evidenced-based psychotherapy is a key driver for stepped care delivery systems. The Improving Access to Psychological Therapies initiative in the UK (IAPT; Layard, 2006) represents a large-scale (national) attempt to systematise stepped care principles into mental health care. Although stepped principles are being increasingly advocated and applied across services, the model itself remains relatively under-evidenced (Bower & Gilbody, 2005). A wide range of uncertainties remain regarding the most clinically and organisationally (i.e. cost and time) effective means of service delivery for depressed patients (Richards & Suckling, 2009). The central objective of this review is therefore to assess evidence for the clinical effectiveness of stepped care interventions for adults with depressive disorders. This will enable service design and redesign to be in line with current evidence.

Method

Study Selection Criteria

The inclusion criteria for studies were as follows: (a) be published in a peer-reviewed journal in the English language, (b) have an experimental or quasi-experimental (i.e. empirical) design, (c) use quantitative outcome measures, (d) evaluate effectiveness in terms of depression-related outcomes, (e) use a working age adult sample,(f) use a sample clinically indicated to have a depressive disorder, (g) have over 50% patients indicated to have a depressive disorder in the sample or subsample reported, and (h) tested a stepped care system. Stepped care systems were defined as the system explicitly using predefined separate intervention components (‘steps’) that increase in intensity and patient burden according to patient need. Patients are stepped up to higher, more intensive steps only on the basis of predefined and explicit criteria (e.g. non-response shown on sessional outcome measures or an increase in risk). Pure stepped care entails starting all patients at

the lowest step and then stepping up accordingly, whilst stratified stepped care entails some patients being automatically allocated to higher steps. This review was not limited to randomised controlled trials (RCTs) as uncontrolled or non-randomized studies were also permitted. This was in order to gather a wider evidence base and to enable the realities of clinical practice to be more closely reflected.

Search Strategy

A systematic search strategy was undertaken. Searches took place in March and April 2013. The following databases were searched; Web of Science, Journal Citation Reports, and BIOSIS citation index and Previews, (via Web of Knowledge), PsychINFO (via OvidSP), and Scopus. Compound search strings targeting titles, topics and keywords were used, incorporating wildcards, Boolean operators and lemmatization where available. Search terms included “stepped care”, “stepped-care”, “stepped * care”, “depression”, “depressive”, “mood”, “distress”, “dysthymia”, and “affective”. After duplicates were removed, titles and abstracts of all unique results returned from databases were screened based on selection criteria (358 studies, k). Full texts of potentially suitable articles were then retrieved and re-examined according to criteria. From these full texts, 14 studies appropriate for inclusion were identified. Searches and study inclusion screenings were undertaken by the lead author, with consultation from the second and third author and an independent consultant clinical psychologist.

Clark et al.’s study (2009) reported results from two separate intervention sites with different samples and analyses. One of these samples (site one) was re-used by Richards & Suckling (2009) and as part of a larger sample in Richards and Borglin’s (2011) study. As the latter comprised the largest overall sample, Richards and Suckling (2009) was excluded, as well as the site one findings from Clark et al. (2009). Clark et al’s (2009) site two findings have been included.

Reference lists of eligible articles were also assessed for inclusion. One additional article was included, making a total of k = 14 articles for review (Figure 1).

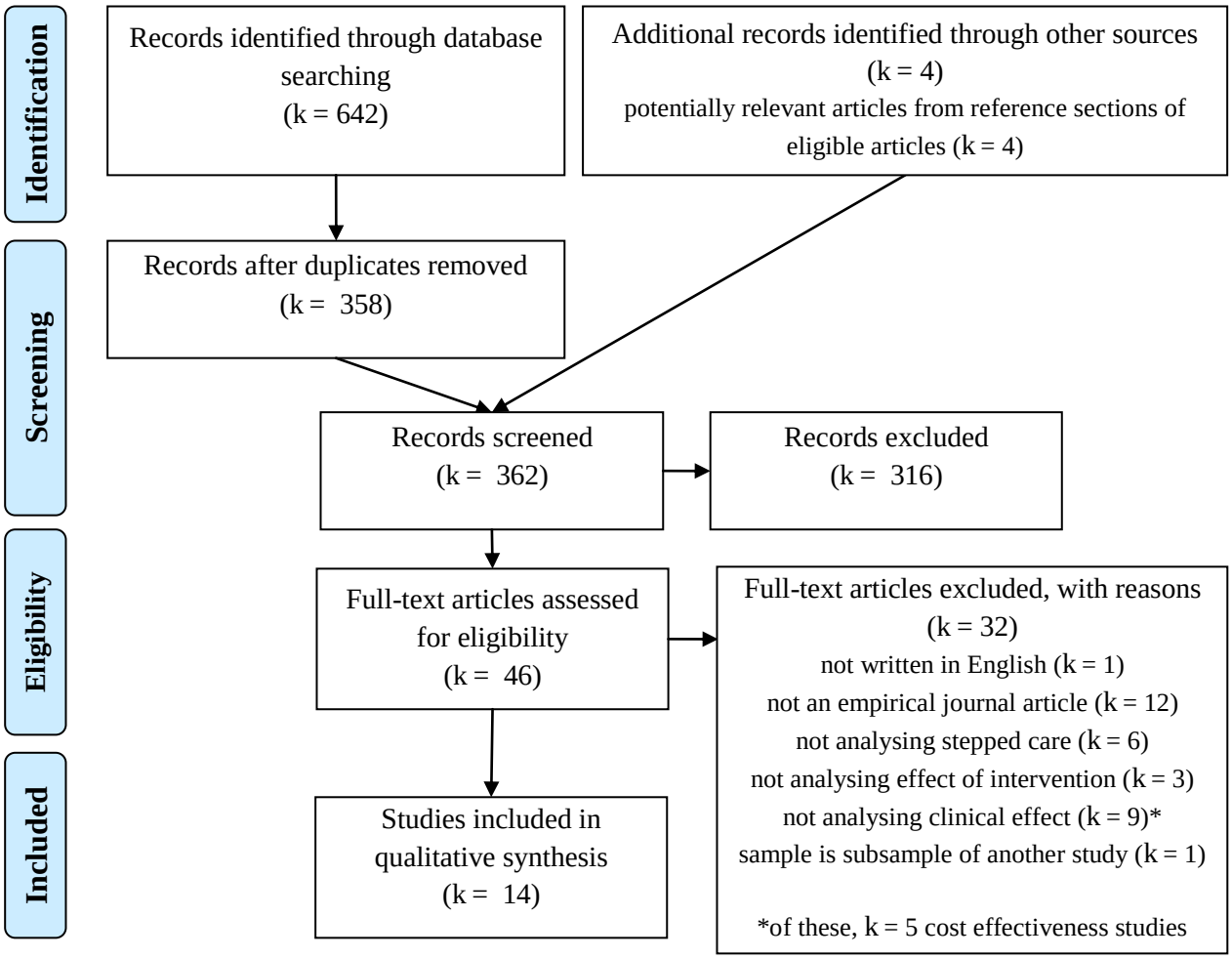

Figure 1. PRISMA diagram indicating flow of records included and excluded from review.

Data Analysis

In line with statistical norms in the majority of published research, the threshold for statistical significance used in this review is an alpha value of 0.05. Confidence intervals reported for odds ratios, risk ratios, and effect sizes are similarly 95% unless otherwise stated. Stepped care approaches to the organisation of care are referred to as intervention systems. Usual care or other care systems are referred to as comparison systems. In cases where intention to treat (ITT) analyses and completer analyses have been reported and results are equivalent, only the ITT analyses have

been reported, as these analyses are considered to have greater clinical significance. Recovery and response rates have been used as the principal measures of system effectiveness. Recovery is defined as scoring below the clinical cut-off on the appropriate standardised measure, whilst response is defined as a 50% reduction in an outcome score. For the majority of the review only depression outcomes are reported. However, a section has been included later to briefly consider other outcomes, thereby enabling the consideration of breadth of outcome. Two studies used considerably larger samples than any other study, skewing the mean study sample size. For this reason, summary statistics relating to sample size have been reported using the median (Mdn) value rather than the mean.

Review

Study Design Synthesis

Table 1 summarises basic study information. The number of patients per study ranged between 18 and 7,859 (Mdn = 430, k = 14). Patient numbers in stepped care systems ranged between 7 and 7,859 patients (Mdn = 204, k = 14), whilst comparison system numbers ranged between 11 and 1,436 patients (Mdn = 126, k = 11). Intervention and comparison systems differed in size because three studies did not include a comparator. Percentage of male participants ranged between 0% and 56% (k = 13) and mean patient age ranged between 35 and 61 (k = 11). Employment ranged between 11% and 66% (k = 7). Employment rates were low in the studies that reported them, although this may have been influenced by selection bias. Causal factors may have included comorbid physical conditions (e.g. cancer in Dwight-Johnson et al., 2005; 11% employment), and socioeconomic factors (e.g. deprivation in Araya et al., 2003; 15% employment). The educational experience of patients could not be appropriately synthesised, but reflected a range of educational levels. Ethnicity and nationality tended to be relatively homogenous within studies, but varied between studies.

Table 1. Summary of Study Characteristics

Lead Author	Year	Notable Co-morbidity	n	RCT	Intervention Steps	Comparison
Studies Conducted Without Comparison Systems						
Clark	2009	Anxiety disorders	1654	NO	1) low intensity, 2) brief, and 3) high intensity CBT-based interventions	-
Franx	2009	-	543	NO	1) psycho-education, self help, counselling, 8 sessions brief psychotherapy, exercise, "other", 2) psycho-education, medication, psychotherapy ("GT, CGT, IPT"[a]), "other".	-
Richards	2011	Anxiety disorders	7859	NO	1) low intensity CBT-based interventions, 2) high intensity CBT-based interventions	-
Studies Conducted With Comparison Systems						
Araya	2003	-	240	YES	1) psychoeducational group and booster sessions, 2a) additional assessment for pharmacotherapy, 2b) refer for primary care physician re-asssessment, initiate or adjust pharmacotherapy	usual care
Davidson	2010	Acute coronary syndrome	237	YES	1) PST or pharmacotherapy, 2) switch treatment, add alternative treatment, or intensify original treatment	usual care
Dwight-Johnson	2005	Cancer	55	NO	1) PST or pharmacotherapy, plus patient information, 2) switch treatment, add alternative treatment, or intensify original treatment	usual care
Ell	2008	Cancer	472	YES	1) PST or pharmacotherapy, plus patient information, 2) switch treatment, add alternative treatment, or intensify original treatment	usual care + pamphlet + resource list
Ell	2010	Diabetes	387	YES	1) PST/pharmacotherapy, maintenance/relapse prevention, 2) switch treatment, add alternative treatment, or intensify original treatment, 3) same as step two, plus potential referral to specialty mental health care.	usual care + pamphlet + resource list
Ell	2011	Diabetes	264	YES		
Kay-Lambkin	2010	Methamphetamine use	18	NO	1) brief integrated CBT/MI intervention (1 session), feedback, self-help and case formulation, 2) +4 sessions, 3) +4 sessions, 4) +4 sessions	all steps of intervention
Patel	2010	Anxiety disorders	774[b]	YES	1) advice, psychoeducation, 2) pharmacotherapy or IPT, adherence management, 3) additional medication or IPT, 4) existing treatment & referral to clinical specialist	usual care + treatment manual
Patel	2011	Anxiety disorders	774[b]	YES		
Seekles	2011	Anxiety disorders	120	YES	1) watchful waiting, 2) guided self help, 3) short face-to-face problem solving, 4) pharmacotherapy and/or specialised MH care	usual care
van Straten	2006	Anxiety disorders	702	YES	1) CBT or brief therapy, 2) pharmacotherapy and/or swap therapy	Matched care

[a] abbreviations not explained in original text. [b] subsample with depression.

CBT = cognitive behavioural therapy, IPT = interpersonal psychotherapy, MI = motivational interviewing, PST = problem-solving therapy, RCT = randomised controlled trial

Diagnostic measures and criteria used by reviewed studies to assess depression are shown in Figure 2. Most criteria used were considered appropriate. Richards and Borglin (2011) stated that lack of standardised diagnostic procedures was a weakness of their study. Franx et al. (2009) only stated that general practitioners were asked to differentiate between severely depressed and non-severely depressed participants. For those studies that either did not state suitable criteria or stated criteria that also included disorders other than depression (Clark et al., 2009; Dwight-Johnson et al., 2005; Ell et al., 2008; Franx et al., 2009; Richards & Borglin, 2011), suitability for review was independently assessed using the percentage of patients meeting either PHQ-9 or BDI clinical cut-offs (both ≥ 10) (Beck et al., 1988; Kroenke et al., 2001; Martin et al., 2006).

Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria	
Composite International Diagnostic Interview for DSM-IV (CIDI) (World Health Organisation, 1990)	Seekles et al. (2011) van Straten et al. (2006)
Mini International Neuropsychiatric Interview (MINI) (Sheehan et al., 1994)	Araya et al. (2003)
International Classification of Diseases (ICD-10) criteria	
Revised Clinical Interview Scale (CIS-R) (Lewis, Pelosi, Araya, & Dunn, 1992)	Patel et al. (2010; 2011)
Condition non-specific diagnostic assessment, based on the ICD-10 framework	Clark et al. (2009)†
Self-report measures	
Beck Depression Inventory (BDI) (Beck, Steer, Ball, & Ranieri, 1996)	Davidson et al. (2010) score ≥10 Kay-Lambkin et al. (2010) score ≥ 17
Patient Health Questionnaire (PHQ-9) score ≥ 10 (Gilbody, Richards, & Barkham, 2007), plus one of two cardinal depression symptoms	Dwight-Johson et al. (2005)‡† Ell et al. (2008)‡† Ell et al. (2010; 2011)
Not clearly stated	Franx et al., (2009)† Richards and Borglin (2011)†
† also independently assessed by first author for inclusion into review, using percentage of participants meeting PHQ-9 or BDI clinical threshold (scores ≥ 10). ‡ study also included participants with dysthymia according to DSM-IV criteria.	

Figure 2. Diagnostic measures and criteria used by reviewed studies.

Components of psychological interventions included brief therapy (BT; Schaefer et al., 1999, as cited in van Straten et al., 2006), low and high intensity cognitive behavioural therapy (CBT; Kuyken et al., 2007), interpersonal psychotherapy (IPT; Klerman et al., 1984), motivational interviewing (MI; Miller & Rollnick, 2012) and problem-solving therapy (PST; Mynors-Wallis et al., 2000). Other intervention components principally included anti-depressant medication (pharmacotherapy), self-help and psychoeducation. Major outcome measures are shown in Figure 3.

- Beck Depression Inventory (BDI) (Beck, Steer, Ball, & Ranieri, 1996)
- Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM) (Barkham et al., 2001)
- Generalised Anxiety Disorder Assessment (GAD-7) (Spitzer, Kroenke, Williams, & Lowe, 2006)
- Hamilton Depression Rating Scale (HDRS) (Hamilton, 1960)
- Hopkins Symptom Checklist (SCL-20) (Derogatis, Rickzels, Uhlenhuth, & Covi, 1974)
- Inventory of Depressive Symptomatology (IDS) (Rush, Gullion, Basco, Jarrett, & Trivedi, 1996)
- Patient Health Questionnaire (PHQ-9) (Cameron et al., 2008; Gilbody, Richards, & Barkham, 2007)
- Short Form 12 and 36 Questionnaires (SF-12 & SF-36) (Ware, Kosinski, & Keller, 1996; Ware, 2000)

Figure 3. Major outcome measures used by reviewed studies.

Nine studies were randomised controlled trials (RCTs; Araya et al., 2003; Davidson et al., 2010; Ell et al., 2008, 2010, 2011; Patel et al., 2010, 2011; Seekles et al., 2011; van Straten et al., 2006). There was one randomised controlled pilot (Dwight-Johnson et al., 2005) and one quasi-randomised comparison study (Kay-Lambkin et al., 2010). The final three studies were uncontrolled prospective cohort studies (Clark et al., 2009; Franx et al., 2009; Richards & Borglin, 2011).

All studies were assessed for quality using the Downs and Black checklist (Downs & Black, 1998). The checklist is suitable for randomised and non-randomised studies, and covers study reporting, external validity, and internal validity. As in other reviews (e.g. Samoocha et al., 2010), the checklist was modified slightly. Item 27 was scored as 0 or 1 (rather than 0 to 5) giving each paper an overall score of 0 to 28. A random sample of three studies was second rated by the third author; 77 of 81 ratings agreed between raters (Cohen's Kappa = 0.87; 95% CI = 0.74 to 0.99). Inconsistencies were resolved and ratings were re-checked. Overall quality ratings for each study are shown in Figure 4 (full item-by-item ratings are available in Appendix A). Using qualitative ranges proposed by Samoocha et al. (2010), all randomised controlled studies were good quality (scores of 20 to 25). Two uncontrolled studies were fair quality (15 to 19; Clark et al., 2009; Richards & Borglin, 2011), and two were poor quality (less than 14; Franx et al., 2009; Kay-Lambkin et al., 2010). No studies were excellent quality (26 to 28).

Figure 4 also shows risk of bias for the randomised controlled studies, according to the Cochrane risk of bias tool (Higgins et al., 2011). This tool is specifically designed to assess quality in RCTs and covers seven areas of bias. Negative areas indicate higher risk of bias. Studies were rated by the lead author according to Cochrane criteria. All 70 ratings were audited by discussion with the second author. From this audit, 20 ratings were challenged and re-assessed, resulting in 7 ratings being adjusted. Full discussion of the quality of evidence is presented in the discussion section.

	Randomised Controlled Studies										Other Studies			
	Araya 2003	Davidson 2010	Dwight-Johnson 2005	Ell 2008	Ell 2010	Ell 2011	Patel 2010	Patel 2011	Seekles 2011	van Straten 2006	Clark 2009	Franx 2009	Kay-Lambkin 2010	Richards 2011
Downs and Black (modified) quality score	23	22	25	23	23	23	23	22	21	22	17	9	7	19
Random Sequence Allocation	+	+	+	+	+	+	+	+	+	+				
Allocation Concealment	+	+	?	?	+	+	?	?	+	+				
Blinding of Participants and Personnel	+	+	+	+	?	?	+	+	+	+				
Blinding of Outcome Assessment	+	+	+	+	?	?	+	+	+	+				
Incomplete Outcome Data	+	-	-	+	+	+	?	?	?	+				
Selective Reporting	+	+	-	-	-	+	+	+	+	+				
Other Bias	+	+	-	+	?	?	+	+	-	-				

Figure 4. Quality ratings of all studies (modified Downs and Black score), plus risk of bias for randomised controlled studies (shown across seven areas according to the Cochrane risk of bias tool). + low risk of bias, - high risk of bias, ? inconclusive risk of bias

Studies Conducted Without Comparison Systems

Three studies did not include a comparator system and are therefore evaluations of the effectiveness of stepped care systems. Of these, two studies investigated the IAPT initiative; Richards and Borglin (2011) evaluated a two-year cohort at "site one," whilst Clark et al. (2009) evaluated a one-year cohort at "site two". One-year cohort results regarding site one from Clark et al. (2009) and Richards and Suckling (2009) were excluded from the main review for reasons

already described, but are briefly described where they reflect outcomes not otherwise captured by Richards and Borglin (2011). Whilst these studies did not utilise comparator systems, patient numbers were considerably more extensive than any other studies included for review. Sessional outcome measures were available for samples of between 1,500-7,000 patients. Treatment at both sites involved low and high intensity CBT-based interventions at the various steps. There were some differences between service designs, with site one having two steps and site two having three steps. Although pharmacotherapy was not part of the stepped care system, between 20-55% of patients were receiving medication during interventions. Clark et al. (2009) found no differences in depression recovery rates between those receiving and not receiving medication. At site one, stepping up decisions were made based on patient progress and discussion with the therapist's supervisor and the patient. Richards and Borglin (2011) found a 43% recovery rate at end of treatment for the Patient Health Questionnaire (PHQ-9), with a pre-post treatment effect size of 1.07. Interventions required approximately three hours of contact over five treatment sessions (generally with three or more by telephone).

Some additional results were reported from Richards and Suckling's (2009) and Clark et al's (2009) respective one-year samples. Richards and Suckling (2009) reported a 55% PHQ-9 end of treatment response rate and Clark et al. (2009) reported a pre-post Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM) effect size of 0.98. Clark et al. (2009) found that although PHQ-9 and CORE-OM scores significantly worsened between treatment completion and follow-up, improvements in PHQ-9 and CORE-OM scores compared with baseline remained significant at both later time points. Clark et al. (2009) reported that stepping up was infrequent at site one – only 4% of those participants still meeting clinical caseness at the end of low intensity therapy were stepped up to high intensity, with many instead being referred for external counselling.

In contrast, at site two the highest (third) step was the most commonly delivered intervention (74%), with treatment lasting seven hours on average (Clark et al., 2009). Stepping up

actions were pre-defined as a lack of improvement after specified durations at each step. Again, Clark et al. (2009) found significant improvements in PHQ-9 and CORE-OM scores, with effect sizes of 1.06 and 1.19 respectively. Improvements were maintained at follow-up. The combined PHQ-9/Generalised Anxiety Disorder Assessment (GAD-7) recovery rate was 55% upon completing treatment, dropping to 42% at follow-up. These studies excluded patients with only one contact from the analyses, meaning these were not intention-to-treat analyses and this may have influenced results.

Franx et al. (2009) reported results from an evaluation of a multi-team uncontrolled two-step system (see Table 1). Stepping decisions were based on symptom chronicity, or severity or lack of response to step one treatment. Beck Depression Inventory (BDI-II) recovery rates indicate that by 3-6 months around 30% of patients had recovered. However, outcome completion rates were low and inconsistent. The stepped model was poorly adhered to with evidence of 22% of patients appropriate for step one and 43% appropriate for step two not receiving the appropriate intervention within the specified time period (one month).

Studies Conducted With Comparison Systems

Eleven studies have compared stepped care with other forms of service delivery (predominantly variations on usual care) and can be considered tests of the efficacy of stepped care. Usual care generally involved treatment from patients' General Practitioners (GPs) and was occasionally "enhanced" with psychoeducational information for physicians or patients. Four of these studies investigated short-term intervention effects (6-12 months) in samples with no comorbid physical health difficulties (Araya et al., 2003; Patel et al., 2010, 2011; Seekles et al., 2011). The studies were conducted in Chile, North America, Goa and The Netherlands. Although Seekles et al.'s (2011) step one was ostensibly watchful waiting, this occurred prior to randomization and so is not appropriate to consider for inclusion. All four studies used

psychoeducation or guided self-help as the first step post-randomization, with the addition of medication and finally referral to other professionals at higher steps. Patel et al. (2010, 2011) and Seekles et al. (2011) included additional psychological therapy at their intermediate steps; IPT and PST. All stepped care models were compared with usual care, although Patel et al. (2010, 2011) compared with usual care enhanced with a treatment manual.

Seekles et al. (2011) found that in both care systems at 6-months after inclusion, Inventory of Depressive Symptomatology (IDS) scores significantly decreased and approximately 50% of patients were recovered from depression or anxiety (Composite International Diagnostic Interview; CIDI). Depression and anxiety recovery rates were not differentiated. No significant outcome differences were found between the care systems (IDS comparative effect size = 0.11 at 6-months). This study was underpowered and suffered high attrition rates. Also, just 58% of participants had depressive disorders, with 86% having comorbid disorders. The remaining 42% of patients had anxiety disorders only and so the relevance of these findings specifically for depression should be treated with due caution. Conversely, Araya et al. (2003) found that stepped care resulted in significantly lower Hamilton Depression Rating Scale (HDRS) and Short Form 36 Questionnaire (SF-36) scores when compared with usual care. At 6-months, the HDRS recovery and response rates were in stepped care were 70% and 78%, compared with usual care rates of 30% and 32%.

Although Patel et al.'s (2010; 2011) overall sample had mixed diagnoses, results reported here specifically relate to their depression sub-sample (n = 774; 35% of the overall sample). Analyses at 6-months were not stratified (Patel et al., 2010), but at 12-months were stratified by public and private facilities (Patel et al., 2011). At 6-months, the stepped care recovery rate from common mental health disorders (Revised Clinical Interview Scale; CIS-R) was 54% - but this was not significantly different from enhanced usual care. The adjusted usual care risk ratio was 1.05 (95% CI = 0.81 to 1.36). At 12-months, stepped care recovery rates were 58% across both public and private facilities. This was significantly better than enhanced usual care in public facilities (42%), but not private facilities (64%). Adjusted risk ratios were 0.76 (95% CI = 0.59 to 0.98) and

1.20 (95% CI = 0.82 to 1.67) respectively. Mean CIS-R symptom score reductions of around 60% were observed in stepped care at 6 and 12-months and in both private and public facilities - but were not significantly different at either time point from enhanced usual care. Finally, there was no statistical difference after 12-months in suicide plans/attempts between the care systems. The authors report that they used an intention to treat analysis (ITT), but describe this as including only those participants seen at 6-months, which would not fit with an ITT approach.

Kay-Lambkin et al.'s (2010) study evaluated a stepped care intervention for patients with depression and methamphetamine use. The stepped care intervention comprised 1-13 sessions of CBT/MI sessions with feedback and self-help. Patients were able to choose the focus of different therapy sessions (depression, methamphetamine use or integrated). This system was compared to a fixed integrated approach, comprising all steps of the stepped care intervention (i.e. all 13 sessions). The sample (n = 29) was the smallest of all reviewed studies and the authors did not therefore attempt statistical analysis. Qualitatively, patients in the stepped care condition reported broadly equivalent BDI-II depression scores to the fixed condition participants over the observed 5-month period.

Five studies investigated the efficacy of stepped care with patients with comorbid physical health conditions. Patients in these samples either had acute coronary syndrome (ACS; Davidson et al., 2010), cancer (Dwight-Johnson et al., 2005; Ell et al., 2008), or diabetes (Ell et al., 2010, 2011). Treatments were similar, all initially involving PST or pharmacology and stepping up involved either substituting treatment, adding the alternative treatment or intensifying the original treatment. Ell et al. (2010, 2011) added a third step to either review treatment again according to step two or to refer to speciality mental health care. Step-up decisions were made based on symptom score reviews every 8-weeks.

Davidson et al.'s (2010) sample excluded those patients whose symptoms spontaneously remitted or responded to usual care within 3-months of ACS. Significantly reduced BDI scores

were found in both stepped and usual care systems at 6-months, but with significantly greater reductions in stepped care compared with usual care. Recovery and response rates were not reported. Dwight-Johnson et al.'s (2005) pilot study found similarly positive results. At 4-8 months, patients in the stepped care system had significantly greater depression (PHQ-9) response rates (37%) than usual care (12%), as well as greater improvement in emotional well-being. However, the pilot was limited by a number of factors, including a small sample size (n = 55). Ell et al.'s (2008) RCT of the same system with a separate sample (n = 472) therefore addressed most of these limitations. At 6-months, neither depression nor emotional well-being outcomes differed significantly between the care systems. However, at 12-months both depression (PHQ-9) response rate (63%) and emotional wellbeing were significantly greater for stepped care. Conversely, change in mean depression (PHQ-9) scores was not significantly different between care systems and Short Form 12 Questionnaire (SF-12) mental health scores showed the reverse pattern, with significance in favour of stepped care disappearing at 12-months. Recovery rates were reported only for those patients completing both 6 and 12-month follow-ups, which might have biased findings. With this caveat, 70% and 73% stepped care depression (PHQ-9) recovery rates were respectively reported at these time points. Enhanced usual care recovery rates were not reported.

In summary, across these findings depression recovery rates appear to vary in stepped care between 50% and 60% and this might be expected within 12-months after initiating treatment. Whilst equivalence to usual care is suggested by comparison studies, clear evidence regarding superiority appears currently inconclusive.

Studies with Long-Term Follow-Up

Although rapid spontaneous remission can be feature of recovery from depression, a chronic disease course is common for those who do not quickly recover (Richards, 2011; Spijker et al., 2002). Evidence however suggests that depression treatment effects can diminish over time (for

example discussion, see Katon et al., 2002) and can sometimes fail to significantly shorten the course of depression (Spijker et al., 2002). Furthermore, rates of recurrence are high even after treatment, especially when sub-threshold symptoms persist (Burcusa & Iacono, 2007; Lin et al., 1998; Richards, 2011). With such factors in mind, it is especially important to consider the durability of outcomes achieved by stepped care service delivery models.

Ell et al. (2010) found that stepped care patients' Hopkins Symptom Checklist (SCL-20) recovery rate rose from 38% to 40% between 6 and 18-months, significantly higher than enhanced usual care at those time points (28% and 35%). Similarly, stepped care SCL-20 response rates (57-62%) were significantly larger at all times than enhanced usual care rates (36-44%). Significant differences in favour of stepped care were also reported for depression (PHQ-9) recovery and response rates. At 24-months, Ell et al. (2011) found recovery rates were matched at 33%, but the adjusted odds ratio remained significant in favour of stepped care. However, the stepped care response rate (58%) was no longer significant in comparison to enhanced usual care (49%). Conversely, stepped care depression (PHQ-9) response rates (54%), but not recovery rates (30%), were significantly better than enhanced usual care at 24-months. Finally, Ell et al. (2010) found that patients receiving treatment in a stepped care system had significantly greater SF-12 mental health scores at 6-18 months compared with enhanced usual care. This significance had disappeared by 24-months post treatment (Ell et al., 2011).

van Straten et al.'s (2006) multi-centre randomised design included follow-ups at 12-months and up to 24-months post-randomisation (18-months, n = 299; 21-months, n = 121; 24-months, n = 64). Differences in follow-up duration were controlled for. The study investigated two different stepped care systems in comparison to "matched care" as usual. Both intervention systems involved CBT, BT and/or medication, but one delivery system began with CBT, whilst the other began with BT. Matched care involved matching therapeutic approach to the patient based on decisions regarding individual needs. Step-up decisions were made on the basis of clinician or patient perceptions of the clinical effectiveness. CIDI recovery rates in the two stepped care arms were

almost identical; both approximated 54% at 12-months and 68% on study completion. No significant differences in outcomes were found between intervention systems or in comparison to matched care. The study lacked power to detect the significance of trends in favour of stepped care that were observed. Despite this, both stepped care systems involved significantly shorter treatment durations (in days) than in matched care.

Synthesis of Depression Effectiveness

Table 2 summarises outcomes. Figure 5 displays comparisons between stepped care systems and comparison systems, showing Cohen's d effect sizes for comparative mean change in outcome scores, and recovery odds ratios. Both are shown over the longest available time period for each sample, and only one outcome measure was used per sample. Estimated values had to be calculated for some studies, using unadjusted available data. In particular, only one comparative value for Cohen's d was available (Seekles et al., 2011; d = 0.11). Four values were estimated from available data (Ell et al., 2008; d = 0.25, Patel et al., 2010; d = 0.41, Davidson et al., 2010; d = 0.45, and Araya et al., 2003; d = 1.09). The median comparative Cohen's d effect size was 0.41 (interquartile intervals 0.25 and 0.45; k = 5). Reported and calculated recovery and response odds ratios for stepped care systems compared with comparison systems were more prevalent, and are shown in Table 2.

Table 2

Summary of Major Depression Related Clinical Outcomes for Stepped Care Interventions

First Author	Year	n	Measure	Re-Assessment	Recovery	Response	Odds Ratio[a] (recovery)	Odds Ratio[a] (response)
Studies Conducted Without Comparison Systems								
Clark	2009	1654	PHQ-9	Treatment end	(NDS) 55% †	.	(n/a)	(n/a)
				+4 – 17 months	(NDS) 42% †	.	(n/a)	(n/a)
Franx	2009	543	BDI-II	3 months	28% †	.	(n/a)	(n/a)
Richards	2011	7859	PHQ-9	Treatment end	43% †	.	(n/a)	(n/a)
Studies Conducted With Comparison Systems								
Araya	2003	240	HDRS	3 months	49% †	55% †	(5.59)	.
				6 months	70% †	81% †	5.52^{o}	7.56^{o}
Davidson	2010	237	BDI	6 months	Not reported	.	.	.
Dwight-Johnson	2005	55	PHQ-9	8 months	.	37% *	.	4.51^{o}
Ell	2008	472	PHQ-9	6 months	70% †	49% $^{-}$	.	1.26^{-}
				12 months	73% †	63% *	.	1.98^{o}
Ell	2010	387	SCL-20	6 months	38% *	57% *	.	2.46^{o}
				12 months	39% $^{-}$	62% *	2.07^{o}	2.59^{o}
				18 months	40% *	62% *	2.66^{o}	2.64^{o}
Ell	2011	264	SCL-20	24 months	33% *	58% $^{-}$	2.06^{o}	1.69^{-}
			PHQ-9	24 months	30% $^{-}$	54% *	1.31^{-}	1.87^{o}
Kay-Lambkin	2010	18	BDI-II	Treatment end	Not reported	.	.	.
Patel	2010	774[b]	CIS-R	6 months	(NDS) 54% $^{-}$	.	(1.14)	.
Patel	2011	774[b]	CIS-R	12 months	(Public Health) (NDS) 58% *	.	(1.90)	.
					(Private Health) (NDS) 58% $^{-}$	.	(0.78)	.
Seekles	2011	120	CIDI	6 months	47% $^{-}$	.	.	.
			IDS	6 months	.	.	(0.86)	.
van Straten	2006	702	CIDI	12 months	(CBT) 53% $^{-}$	.	1.36^{-}	.
					(BT) 55% $^{-}$	.	1.48^{-}	.
				18 – 24 months	(CBT) 67% $^{-}$	.	1.26^{-}	.
					(BT) 69% $^{-}$	.	1.41^{-}	.

All significant results are in favour of the intervention group.

[a] all odds ratios are relative to comparison system, [b] subsample with depression, (italicised parentheses) = odds ratio calculated with available data.

BDI = Beck Depression Inventory, BT = Brief Therapy group, CBT = CBT intervention group, CIDI = Composite International Diagnostic Interview, CIS-R = Revised Clinical Interview Scale, HDRS = Hamilton Depression Rating Scale, IDS = inventory of depressive symptomatology, NDS = non-depression-specific recovery, PHQ-9 = Patient Health Questionnaire, SCL-20 = Hopkins Symptom Checklist, WSAS = Work and Social Adjustment Scale

$^{-}p > .05$, * $p \leq .05$, ^{o} 95% confidence interval not overlapping 1.0, † not compared with comparison group

The median odds ratio for recovery was 1.31 (interquartile intervals 1.05 and 1.66; k = 7). The median odds ratio for treatment response was 3.25 (interquartile intervals 1.98 and 4.51; k = 4). It is important to recognise that values were only obtainable from a relatively small number of included studies. Two studies reported uncontrolled (pre-post) Cohen's d values for stepped care systems; Richards & Borglin (2011) (d = 1.07), and Clark et al., (2009) (d = 1.06).

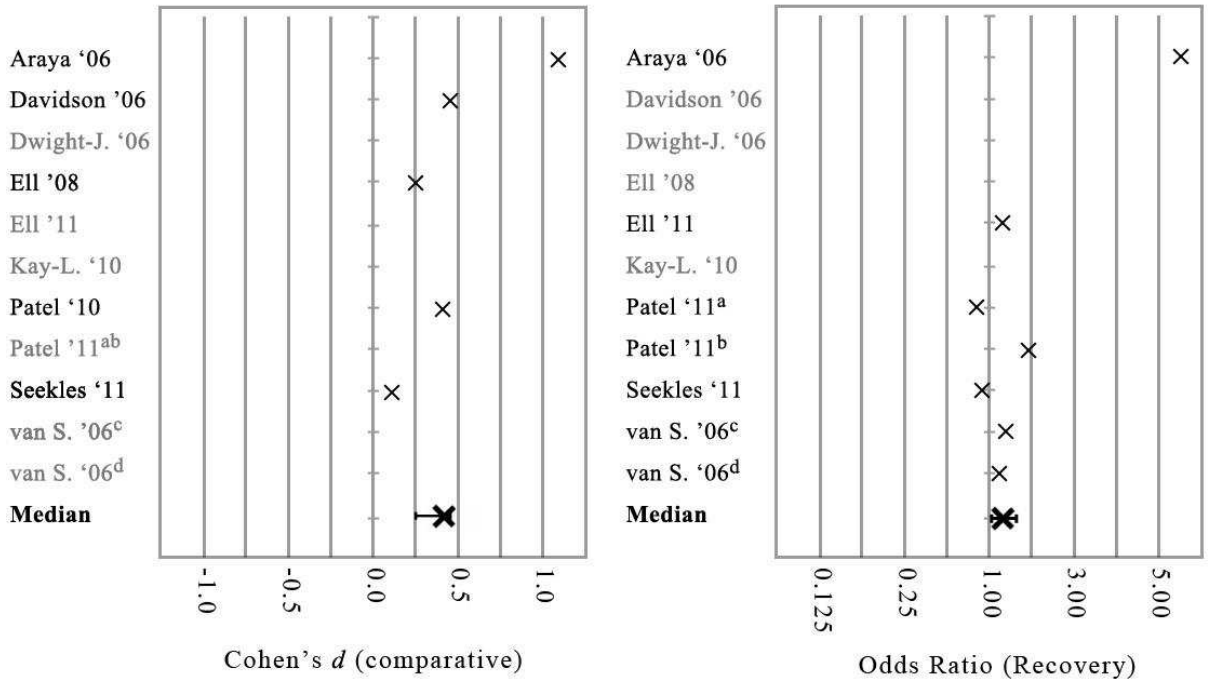

Figure 5. Plot of Cohen's d effect sizes and recovery odds ratios, for stepped care systems compared with comparison systems. Median effect sizes are shown in bold, with error bars indicating interquartile intervals. Effect sizes greater than 0.0 and odds ratios greater than 1.0 indicate stepped care was more effective than the comparison system, and vice versa. Where outcomes from the same sample were recorded at more than one time point (including in more than one study), the longest time point was used. Values for studies in grey were not reported and were unable to be calculated using available data.

[a] private care subsample. [b] public care subsample. [c] brief therapy subsample. [d] cognitive behavioural therapy subsample.

Findings from the reviewed studies give some indication that positive clinical outcomes in stepped care systems can be maintained for long periods, but that some of the benefits over care as usual are lost over time. However, there is insufficient evidence to draw firm conclusions.

It also appears that stepped care interventions can be effective in populations with comorbid physical and mental health conditions. In particular, seven studies investigated samples with notable proportions of patients with comorbid anxiety disorders (more than 25% of study sample; see Table 1). Although these proportions were reported by studies in different ways and cannot easily be meaningfully synthesised, in general, the proportion of patients with anxiety disorders varied between approximately 40% and 70%. All seven studies reported reductions in depression symptoms, with 30-60% recovery rates. However, of the four studies which compared stepped care with usual or matched care (Patel et al., 2010, 2011; Seekles et al., 2011; van Straten et al., 2006), only Patel et al. (2011) found significant benefits of stepped care.

Other Non-Depression Outcomes Reported

Richards and Borglin (2011) reported that in relation to anxiety, 40% of patients recovered and reliably improved (GAD-7), with a further 15% reliably improving (effect size 1.04; 0.88 – 1.23). Both Clark et al. (2009) and Seekles et al. (2011) found significant reductions in anxiety scores, although Seekles et al. (2011) found no significant difference between stepped and usual care. Conversely, Ell et al. (2011) found that patients receiving therapy in stepped care systems were less anxious than enhanced usual care at 6, 12, and 24 months. Clark et al. (2009) reported that a significant number of patients returned to work from statutory sick pay (around 10% return to work rate at both sites). Davidson et al. (2010) found that patients treated with stepped care systems reported significantly fewer non-depression related psychiatric problems and higher survival rates for major adverse cardiac events. Qualitative findings from Kay-Lambkin et al.'s (2010) system comparison suggested halving of methamphetamine use, although this was comparable to change in

the comparison group and was not statistically analysed. Conversely, in other studies no significant differences were found between systems on measures of diabetes self-management, WHODAS disability scores, or disability days taken by participants (Ell et al., 2011; Patel et al., 2011).

Ell et al. (2008) and Araya et al. (2003) found significant improvement effects in favour of stepped care on a range of physical, social and emotional quality of life scores including subscales of the SF-12. Similarly, Ell et al. (2010; 2011) found significantly greater reduction in SF-12 physical component scores, functional impairment (SDS) scores, pain impact and diabetes symptoms at 6-months compared to enhanced usual care. However, these effects effectively disappeared at 12-months.

Discussion

Quality Critique

The majority of studies randomized patients to care systems, but three studies did not include any comparator system(s). Although blinding patients and therapists to care system was difficult to achieve, blinding of researchers was possible and yet less commonly completed. One fifth of the studies used particularly small sample sizes (Dwight-Johnson et al., 2005; Kay-Lambkin et al., 2010; Seekles et al., 2011). Power analyses were reported by eight studies, of which two were underpowered (Seekles et al., 2011; van Straten et al., 2006). Conversely, many studies included suitable methods for handling or imputing missing data and ran multiple comparative analyses (for example completers only compared with ITT or sensitivity analyses). A third of studies specifically aimed to reflect routine practice, but such studies tended to be underpowered or suffered other serious methodological difficulties. A number of studies did not report clinical outcome measure means or SDs and the type of effect size reported varied between studies, making comparisons or meta-analysis more difficult.

Acceptability of Intervention

Effectiveness for patients is contingent, in part, on system acceptability. Three studies (Davidson et al., 2010; Ell et al., 2010, 2011) assessed acceptability and reported significantly greater satisfaction with stepped care compared with traditional systems. Some authors included patient choice as a criterion for stepping treatment up or down. Seekles et al. (2011) and Kay-Lambkin et al. (2010) reported that most unplanned attrition from treatment occurred during the initial low-intensity step, suggesting that failure to respond to the initial step may actually discourage patients from further engagement across higher steps. Both advocated expanded patient choice, for example, regarding access to different interventions within steps or the intensity of intervention across steps (e.g. timing, frequency and style of depression treatment sessions). This raises the key dilemma in stepped care systems of balancing the efficient distribution and composition of organisational resources across various steps and the importance of access to a choice of effective and comprehensive treatments in the early steps.

Intervention Effectiveness

Depression recovery rates were predominantly between 40-60%, with depression response rates of around 60%. For stepped care compared with comparison systems, the median recovery odds ratio was 1.31 (interquartile intervals 1.05 to 1.66; k = 7), and the median response odds ratio was 3.25 (interquartile intervals 1.98 and 4.51; k = 4). The median comparative Cohen's d effect size was 0.41 (interquartile intervals 0.25 and 0.45; k = 5). Although stepped care response rates were in general significantly higher than in comparison systems, half of the studies that compared stepped care recovery rate with a comparison system found no significant differences between the systems. There was insufficient information to draw conclusions regarding effectiveness based on treatment duration, step-up rates or similar characteristics. Tentative categorisation of studies by

treatment modality suggested broadly similar outcomes irrespective of the treatment model (i.e. CBT-based interventions, PST and pharmacology-based interventions and 3-step systems involving psychoeducation, psychological therapy/pharmacology and referral). Patient severity and symptom chronicity varied considerably, and there were no clear trends that related chronicity/severity to clinical outcome. Unfortunately, a number of studies either investigated samples in which not all patients had depression or failed to differentiate depression and anxiety outcomes. Such outcomes may be confounded to some extent. Despite this, stepped care appears to be resistant to cultural variation and across samples with various comorbid difficulties.

In the current review, evidence comparing the clinical effects of stepped care with usual care was mixed and one reason for this may be study and system heterogeneity. There were three main identified sources of heterogeneity. Firstly, sample demographics varied considerably as outlined above. Secondly, there was marked variation in the effectiveness of "usual" care. Stepped care was found to be more effective than Goan public (but not private) health care (Patel et al., 2010). In contrast, none of the three Dutch studies found significant differences between care systems, despite stepped care recovery rates of up to 70% (van Straten et al., 2006). Seekles et al. (2011) suggest that Dutch health services are already developed and so stepped care might struggle to offer added value or clinical efficacy. Thirdly and crucially, it is important to recognise the extent of heterogeneity between the stepped care systems themselves.

System Heterogeneity

There was considerable heterogeneity between stepped care systems, including wide variation in the components of treatment and in the criteria/timing of stepping-up processes. For example, whilst some stepped care systems had psychological therapy as a final step, others employed it at step one. Medication was sometimes included as an explicit step (or steps), but sometimes independently managed by other professionals. Some systems involved receiving either

intervention X or Y at step one, before stepping "up" to the alternative intervention at step two. In other systems stepping up involved increasing the duration or intensity of a specific treatment (e.g. number of sessions or medication dosage). Finally, a number of systems did not discuss any operationalisation of stratified stepped care and so had no apparent provision for starting patients on a higher step (or skipping steps), instead requiring patients to progress through each step. As stepping up normally is based on lack of clinical effectiveness of the previous step, then the impact of such treatment failure episodes on patient depression schema and mood need to be considered. Although variations in system implementation are understandable given the varying needs of different clinical populations, many of the stepping decisions in the studies did not appear to be grounded in any particular evidence (or were not explained, if so).

These issues highlight the question of what exactly a stepped care system involves, as in practice there remain clear differences in interpretation and associated delivery. For example, Richards and Suckling (2009) argue that the balance of stepped care rests "between stepped models (where almost all patients are allocated to lower steps initially) and stratified models (where patients are allocated to steps using criteria applied at assessment)". Trials need to be conducted comparing the clinical and organisational efficiency of 'pure' stepped versus more stratified care models, as well as potential component analyses within systems.

Specificity of Effect

One difficulty with the studies reviewed is that the magnitude of effect specifically due to stepped care is unclear and/or hard to isolate. Although all studies used stepped care systems, the impact of implementing stepped care may have been confounded or occluded by other effects. For example, six studies reported using systems that incorporated and mixed both stepped care and collaborative care principles (Dwight-Johnson et al., 2005; Ell et al., 2008, 2010, 2011; Patel et al., 2010, 2011). It is difficult to determine what proportion of effect is related to the stepped aspect,

compared with the collaborative care aspects. This is particularly important given these studies were the only ones other than Araya et al. (2003) to detect significantly better recovery or response rates than comparator systems.

Furthermore, almost all studies compared stepped systems with usual care. It is unclear (and arguably unlikely) that all the intervention options available in stepped care systems were also available in usual care systems. This means the effect of the interventions may be due to a specific component or components of therapy or service delivery, rather than the stepped care framework itself. In short, the current literature is insufficiently controlled and so more methodologically robust research on stepped care service configurations is essential. Ideally, comparison of different but similar stepped care models is needed to identify active/important factors and their ideal specification (Richards & Borglin, 2011). Stepped care component analyses are again indicated.

Clinical Implications

Stepped care system designs appear to be effective for treating depression, although the specific active components are currently unclear. More research is needed to determine whether stepped care systems are more efficient and efficacious than usual care. It should be acknowledged that this depends on both the quality of the comparator and stepped care system, so that like is being compared with like. Provision to allow more severely depressed patients to be stepped up appropriately are important to ensure that clinical guidelines are followed and this was not always evident in reviewed studies. Additionally, assigning patients to low steps without sufficient explanation or patient collaboration could alienate patients and impact on future engagement and intervention acceptability.

Limitations of this Review

The decision to include non-RCT studies may have increased the frequency of methodological weaknesses in the included studies and could be seen as a threat to the overall quality of the evidence base. A meta-analysis of clinical outcomes has not been included in this review and would also have been desirable. This was considered unrealistic, given that (a) reported outcome measures and statistics varied considerably between different studies, and (b) in many cases, suitable outcomes unable to be calculated given the available data. Finally, the majority of the review itself was conducted by the lead author, with consultation from the second and third authors (for example, discussion of studies where inclusion/exclusion was unclear). Independent or collaborative review, or sustained collaboration with other leading experts might have provided a further check for accuracy and served to bring new insights and perspectives to the review.

Conclusions and Recommendations

In conclusion, evidence suggested that stepped care systems were as effective as usual care systems in a range of contexts and with varying populations. Although some studies found that stepped care was significantly more effective than usual care or enhanced usual care, findings overall were inconclusive. Recovery rates according to various measures ranged predominantly between 40-60%, with response rates of around 60%. The quality of evidence was mixed, although studies included a number of well controlled and randomised trials. Specificity of system effect was unclear and it would be helpful for future controlled research to investigate and evaluate specific elements and components of stepped care in order to improve understanding. Further consideration of patients' roles in stepping decisions is also recommended, as well as acceptability of the systems across stakeholders. Samples in the reviewed studies varied considerably and frequently experienced comorbid physical or mental health conditions, supporting the generalisability of findings. The future research agenda therefore suggests benchmarking the clinical and

organisational efficacy of pure stepped care systems versus stratified stepped care versus standard care in terms of costs, multi-stakeholder satisfaction and clinical outcomes all across the short, medium and long-term. A review of the cost-effectiveness of stepped care for depression is particularly recommended. More research isolating and investigating the specifics of stepped service configurations is also encouraged, in order to improve understanding of the active ingredients of stepped care and to inform future stepped care designs in practice.

References[1]

*Araya, R., Rojas, G., Fritsch, R., Gaete, J., Rojas, M., Simon, G., Peters, T.J., 2003. Treating depression in primary care in low-income women in Santiago, Chile: a randomised controlled trial. Lancet 361, 995-1000.

Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., Benson, L., Connell, J., Audin, K., McGrath, G., 2001. Service profiling and outcomes benchmarking using the CORE-OM: Toward practice-based evidence in the psychological therapies. J. Consult. Clin. Psychol. 69, 184-196.

Beck, A.T., Steer, R.A., Ball, R., Ranieri, W. 1996. Comparison of Beck Depression Inventories –IA and –II in psychiatric outpatients. J. Pers. Assess. 67, 588-597.

Beck, A.T., Steer, R.A., Garbin, M.G.J., 1988. Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. Clin. Psychol. Rev. 8, 77-100.

Bower, P., Gilbody, S., 2005. Stepped care in psychological therapies: access, effectiveness and effeiciency. Narrative literature review. Br. J. Psychiatry 186, 11-17.

Burcusa, S.L., Iacono, W.G., 2007. Risk for recurrence in depression. Clin. Psychol. Rev. 27, 959-985.

Cameron, I.M., Crawford, J.R., Lawton, K., Reid, I. C., 2008. Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. Br. J. Gen. Pract. 58, 32-36.

Centre for Economic Performance, 2006. The depression report: A new deal for depression and anxiety disorders. London School of Economics and Political Science Centre for Economic Performance, London.

[1] Papers marked with an asterisk (*) were included in the review

Chatterjee, S., Chowdhary, N., Pednekar, S., Cohen, A., Andrew, G., Araya, R., Simon, G., King, M., Telles, S., Weiss, H., Verdeli, H., Clougherty, K., Kirkwood, B., Patel, V., 2008. Integrating evidence-based treatments for common mental disorders in routine primary care: feasibility and acceptability of the MANAS intervention in Goa, India.World Psychiatry 7, 39-46.

*Clark, D.M., Layard, R., Smithies, R., Richards, D.A., Suckling, R., Wright, B., 2009. Improving access to psychological therapy: Initial evaluation of two UK demonstration sites.Behav. Res. Ther. 47, 910-920.

Cuijpers, P., Smit, F., Oostenbrink, J., de Graaf, R., Ten Have, M., Beekman, A., 2007. Economic costs of minor depression: a population-based study. Acta. Psychiatr. Scand. 115, 229-236.

*Davidson, K.W., Rieckmann, N., Clemow, L., Schwartz, J.E., Shimbo, D., Medina, V., Albanese, G., Kronish, I., Hegel, M., Burg, M.M., 2010. Enhanced depression care for patients with acute coronary syndrome and persistent depressive symptoms: Coronary psychosocial evaluation studies randomized controlled trial. Arch. Intern. Med. 170, 600-608.

Davison, G., 2000. Stepped care: Doing more with less? J. Consult. Clin. Psychol. 68, 580-585.

Derogatis, L.R., Lipman, R.S., Rickels, K., Uhlenhuth, E.H., Covi, L., 1974. The Hopkins symptom checklist (HSCL). A measure of primary symptom dimensions. Mod. Probl. Pharmacopsychiatry 7, 79-110.

Downs, S.H., Black, N., 1998. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J. Epidemiol. Community Health, 52, 377-384.

*Dwight-Johnson, M., Ell, K., Lee, P.J., 2005. Can collaborative care address the needs of low-income Latinas with comorbid depression and cancer? Results from a randomized pilot study. Psychosomatics 46, 224-232.

*Ell, K., Katon, W., Xie, B., Lee, P.J., Kapetanovic, S., Guterman, J., Chou, C.P., 2010. Collaborative care management of major depression among low-income predominantly hispanic subjects with diabetes: A randomized controlled trial. Diabetes Care 33, 706-713.

*Ell, K., Katon, W., Xie, B., Lee, P.J., Kapetanovic, S., Guterman, J., Chou, C.P., 2011. One-year postcollaborative depression care trial outcomes among predominantly Hispanic diabetes safety net patients. Gen. Hosp. Psychiatry 33, 436-442.

*Ell, K., Xie, B., Quon, B., Quinn, D.I., Dwight-Johnson, M., Lee, P.J., 2008. Randomized controlled trial of collaborative care management of depression among low-income patients with cancer. J. Clin. Oncol. 26, 4488-4496.

*Franx, G., Meeuwissen, J.A.C., Sinnema, H., Spijker, J., Huyser, J., Wensing, M., de Lange, J., 2009. Quality improvement in depression care in the Netherlands: The depression breakthrough collaborative. A quality improvement report. International Journal of Integrated Care 9, e84.

Gilbody, S., Richards, D., Barkham, M., 2007. Diagnosing depression in primary care using self completed instruments: a UK validation of the PHQ-9 and CORE-OM. Br. J. Gen. Pract. 57, 650-652.

Haaga, D.A.F., 2000. Introduction to the special section on stepped care models in psychotherapy. J. Consult. Clin. Psychol. 68, 547-548.

Higgins, J.P.T., Altman, D.G., Gøtzsche, P.C., Jüni, P., Moher, D., Oxman, A.D., Savović, J., Schulz, K.F., Weeks, L., Sterne, A.C., 2011. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 343, d5928.

Hunter, D.J., Fairfield, G., 1997. Managed care: Disease management. BMJ, 315, 50-53.

Jaehne, A., Loessl, B., Frick, K., Berner, M., Hulse, G., Balmford, J., 2012. The efficacy of stepped care models involving psychosocial treatment of alcohol use disorders and nicotine dependence: a systematic review of the literature. Current Drug Abuse Reviews 5, 41-51.

Katon, W., Russo, J., Von Korff, M., Lin, E., Simon, G., Bush, T., Ludman, E., Walker, E., 2002. Long-term effects of a collaborative care intervention in persistently depressed primary care patients. J. Gen. Intern. Med. 17, 741-748.

Katon, W., Von Korff, M., Lin, E., Simon, G., Walker, E., Bush, T., Ludman, E.,1997. Collaborative management to achieve depression treatment guidelines. J. Clin. Psychiatry 58(Suppl. 1), 20-23.

*Kay-Lambkin, F.J., Baker, A.L., McKetin, R., Lee, N., 2010. Stepping through treatment: Reflections on an adaptive treatment strategy among methamphetamine users with depression. Drug and Alcohol Review 29, 475-482.

Kessler, R.C., Berglund, P.A., Demler, O., Jin, R., Koretz, D., Merikangas, K.R., Rush, A.J., Walters, E.E., Wang, P.S., 2003. The epidemiology of major depressive disorder: Results from the National Comorbidity Survey replication (NCS-R). JAMA 289, 3095-3105.

Klerman, G.L., Weissman, M.M., Rounsaville, B.J., Chevron, E.S., 1984. Interpersonal Psychotherapy of Depression. Basic Books, New York, NY.

Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The PHQ-9 – Validity of a brief depression severity measure. J. Gen. Intern. Med. 16, 606-613.

Kuyken, W., Dalgleish, T., Holden, E.R., 2007. Advances in cognitive-behavioural therapy for unipolar depression. Can. J. Psychiatry 52, 5-13.

Layard, R., 2006. The case for psychological treatment centres. BMJ, 33, 1030-1032.

Lewis, G., Pelosi, A., Araya, R., Dunn, G., 1992. Measuring psychiatric disorder in the community: a standardized assessment for use by lay interviewers. Psychol. Med. 22, 465-486.

Lin, E.H.B., Katon, W.J., VonKorff, M., Russo, J.E., Simon, G.E., Bush, T.M., Rutter, C.M., Walker,E.A., Ludman, E., 1998. Relapse of depression in primary care – Rate and clinical predictors. Arch. Fam. Med. 7, 443-449.

Martin, A., Rief, W., Klaiberg, A., Braehler, E., 2006. Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. Gen Hosp Psychiatry 28, 71-77.

Miller, W.R., Rollnick, S., 2012. Motivational Interviewing, Third Edition: Helping People Change. Guilford Press, London.

Mynors-Wallis, L.M., Gath, D.H., Day, A., Baker, F., 2000. Randomised controlled trial of problem solving treatment, antidepressant medication, and combined treatment for major depression in primary care, BMJ 320, 26-30.

National Institute for Health and Clinical Excellence, 2005. Obsessive-compulsive disorder: core interventions in the treatment of obsessive-compulsive disorder and body dysmorphic disorder. NICE, London.

National Institute for Health and Clinical Excellence, 2009. Depression: management of depression in primary and secondary healthcare. NICE, London.

National Institute for Health and Clinical Excellence, 2011. Generalised anxiety disorder and panic (with or without agoraphobia) in adults. NICE, London.

Patel, V., 2007. Mental health in low- and middle-income countries. Br. Med. Bull. 81-82, 81-96.

*Patel, V., Weiss, H.A., Chowdhary, N., Naik, S., Pednekar, S., Chatterjee, S., De Silva, M.J., Bhat, B., Araya, R., King, M., Simon, G., Verdeli, H., Kirkwood, B.R., 2010. Effectiveness of an

intervention led by lay health counsellors for depressive and anxiety disorders in primary care in Goa, India (MANAS): a cluster randomised controlled trial. Lancet 376, 2086-2095.

*Patel, V., Weiss, H.A., Chowdhary, N., Naik, S., Pednekar, S., Chatterjee, S., Bhat, B., Araya, R., King, M., Simon, G., Verdeli, H., Kirkwood, B.R., 2011. Lay health worker led intervention for depressive and anxiety disorders in India: impact on clinical and disability outcomes over 12 months. Br. J. Psychiatry 199, 459-466.

Richards, D., 2011. Prevalence and clinical course of depression: A review. Clin. Psychol. Rev.31, 1117-1125.

*Richards, D.A., Borglin, G., 2011. Implementation of psychological therapies for anxiety and depression in routine practice: Two year prospective cohort study. J. Affect. Disord. 133, 51-60.

Richards, D.A., Suckling, R., 2009. Improving access to psychological therapies: Phase IV prospective cohort study. Br. J. Clin. Psychol. 48, 377-396.

Rollman, B.L., Weinreb, L., Korsen, N., Schulberg, H.C., 2006. Implementation of guideline-based care for depression in primary care. Adm. Policy Ment. Health 33, 43-53.

Rush, A.J., Gullion, C.M., Basco, M.R., Jarrett, R.B., Trivedi, M.H., 1996. The Inventory of Depressive Symptomatology (IDS): Psychometric properties. Psychol. Med. 26, 477-486.

Samoocha, D., Bruinvels, D.J., Elbers, N.A., Anema, J.R., van der Beek, A.J., 2010. Effectiveness of web-based interventions on patient empowerment: A systematic review and meta-analysis. Journal of Medical Internet Research 12, e23.

*Seekles, W., van Straten, A., Beekman, A., van Marwijk, H., Cuijpers, P., 2011. Stepped care treatment for depression and anxiety in primary care. A randomized controlled trial. Trials 12, 171.

Sheehan, D.V., LeCrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., 1994. The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J. Clin. Psychiatry 59(Suppl. 20), 20-33.

Siddiqi, K., Siddiqi, N., 2007. Treatment of common mental disorders in primary care in low- and middle-income countries. Trans R Soc Trop Med Hyg 101, 957-958.

Simon, G., 2006. Collaborative care for depression. BMJ 332, 249-250.

Simon, G., Revicki, D., Von Korff, M., 1993. Telephone assessment of depression severity.J. Psychiatr. Res. 27, 247-252.

Singleton, N., Bumpstead, R., O'Brien, M., Lee, A., Meltzer, H., 2001. Psychiatric Morbidity among Adults living in Private Households, 2000. TSO, London.

Sobell, M., Sobell, L., 2000. Stepped care as a heuristic approach to the treatment of alcohol problems. J. Consult. Clin. Psychol. 68, 573-579.

Spijker, J., De Graaf, R., Bijl, R.V., Beekman, A.T.F., Ormel, J., Nolen, W.A., 2002. Duration of major depressive episodes in the general population: results from the The Netherlands Mental Health Survey and Incidence Study (NEMESIS). Br. J. Psychiatry 181, 208-213.

Spitzer, R.L., Kroenke, K., Williams, J.B.W., Lowe, B., 2006. A brief measure for assessing generalized anxiety disorder: The GAD-7. Arch. Intern. Med. 38, 677-688.

van der Leeden, A.J., van Widenfelt, B.M., van der Leeden, R., Liber, J.M., Utens, E.M., Treffers, P.D., 2011. Stepped care cognitive behavioural therapy for children with anxiety disorders: a new treatment approach. Behavioural and Cognitive Psychotherapy 39, 55-75.

*van Straten, A., Tiemens, B., Hakkaart, L., Nolen, W.A., Donker, M.C.H., 2006. Acta. Psychiatr. Scand. 113, 468-476.

van't Veer-Tazelaar, P.J., van Marwijk, H.W., van Oppen, P., van Hout, H.P., van der Horst, H.E., Cuijpers, P., Smit, F., Beekman, A. T., 2009. Stepped-care prevention of anxiety and depression in late life: a randomized controlled trial. Arch. Gen. Psychiatry 66, 297-304.

Von Korff, M., Ormel, J., Katon, W., Lin, E.H.B., 1993. Disability and depression among high utilizers of health care: a longitudinal analysis. Arch. Gen. Psychiatry 2, 91-100.

Wang, P.S., Aguilar-Gaxiola, S., Alonso, J., Angermeyer, M.C., Borges, G., Bromet, E. J., Bruffaerts, R., de Girolamo, G., de Graaf, R., Gureje, O., Haro, J.M., Karam, E.G., Kessler, R.C., Kovess, V., Lane, M.C., Lee, S., Levinson, D., Ono, Y., Petukhova, M., Posada-Villa, J., Seedat, S., Wells, J.E., 2007. Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the WHO world mental health surveys. Lancet 370, 841-850.

Ware, J.E., Jr., 2000. SF-36 health survey update. Spine 25, 3130-3139.

Ware, J.E., Jr., Kosinski, M., Keller, S.D., 1996. A 12-item short-form health survey – Construction of scales and preliminary tests of reliability and validity. Med. Care 34, 220-233.

Wells, K.B., Steward, A., Hays, R.D., Burnam, M.A., Rogers, W., Daniels, M., Berry, S., Greenfield, S., Ware, J., 1989. The functioning and well-being of depressed patients. Results from the medical outcomes study. JAMA 262, 914-919.

Williams, J.B.W., Gibbon, M., First, M.B., Spitzer, R.L., Davies, M., Borus, J., Howes, M.J., Kane, J., Pope, H.G., Jr., Rounsaville, B., Wittchen, H., 1992. The Structured Clinical Interview for DSM-III-R (SCID) II. multisite test-retest reliability. Arch. Gen. Psychiatry 49, 630-636.

Wilson, G., Vitousek, K., Loeb, K., 2000. Stepped care treatment for eating disorders. J. Consult. Clin. Psychol. 68, 564-572.

World Health Organisation, 1990. Composite International Diagnostic Interview (CIDI). World Health Organisation, Geneva.

World Health Organisation, 2001. The world health report 2001 - mental health: New understanding, new hope. World Health Organisation, Geneva.