Appendix A

Study	Organization (practices)	n	Study Quality Analysis (follow-up rate)	Blinding outcomes assessment	Risk of bias
Adler et al., 2004	Academic and community (9 practices; 53 clinicians)	507	Intent-to-treat (3 months, 73%; 6 months, 76%)	Yes	Low
Araya et al., 2003	3 Clinics, Santiago, Chile (clinicians, NS)	240	Intent-to-treat (3 months, 88%; 6 months, 88%)	Yes	Low
Brook et al., 2005	General practice, The Netherlands (19 community pharmacists; clinicians not involved)	135	Intent-to-treat (3 months, 88%; 6 months, 81%)	Yes	Moderate
Bruce et al., 2004	(20 Community practices; number of clinicians, NS)	598	Intent-to-treat (4 months, 82%; 8 months, 76%; 12 months, 69%)	Yes	Low
Capoccia et al., 2004	Academic family practice clinic (20 staff physicians; 18 trainees; 4 Physician Assistants)	74	Intent-to-treat (3 months, 96%; 6 months, 95%; 9 months, 95%; 12 months, 93%)	Yes	Low
Datto et al., 2003	University-affiliated (35 practices; 151 clinicians)	61	Completers (6 months, 86%)	No	High
Dietrich et al., 2004	5 Health care organizations (60 practices; 226 clinicians)	405	Intent-to-treat (4 months, 82%; 8 months, 76%; 12 months, 69%)	Yes	Low
Dobscha et al., 2006	Veterans Affairs (5 practices; 41 clinicians)	375	Intent-to-treat (6 months, 84%; 12 months, 85%)	Yes	Low
Finley et al., 2003	Staff model HMO (1 practice; 30 clinicians)	125	Completers (6 months, 67%)	Yes	High
Fortney et al., 2006	Veterans Affairs: community clinics (7 clinics)	395	Intent-to-treat (6 months, 91%; 12 months, 85%)	Partially met	Moderate
Hedrick et al., 2003	Veterans Affairs (4 firms; 89 clinicians)	354	Intent-to-treat (3 months, 92%; 9 months, 92%)	Yes	Low

			Study Quality		
Study	Organization (practices)	n	Analysis (follow-up rate)	Blinding outcomes assessment	Risk of bias
Hunkeler et al., 2000	Staff model HMO (2 practices; 100 clinicians)	302	Completers (6 weeks, 90%; 6 months, 85%)	NS	Moderate
Katon et al., 1995	Staff model HMO (1 practice; 22 clinicians)	217	Completers (4 months, 89%; 7 months, 85%)	Yes	Moderate
Katon et al., 1996	Staff model HMO (1 practice; 22 clinicians)	153	Intent-to-treat (4 months, 84%; 7 months, 76%)	Yes	Low
Katon et al., 1999	Staff model HMO (4 practices; 73 clinicians)	228	Intent-to-treat (3 months, 85%; 6 months, 84%; 28 months, 75%)	Yes	Low
Katon et al., 2004	Staff model HMO (9 practices)	329	(3 months, 93%; 6 months, 89%; 12 months, 88%)	Yes	Low
Katzelnick et al., 2000	3 HMOs (practices, NS; 163 clinicians)	407	Intent-to-treat (3 months, 94%; 6 months, 94%; 12 months, 93%)	Yes	Low
Mann et al., 1998	General practice, (19 practices)	419	Intent-to-treat (4 months, 92%)	No	High
Oslin et al., 2003	Veterans Affairs (4 practices; 37 clinicians)	97	Intent-to-treat (4 months, 76%)	Yes	Low
Peveler et al., 1999	General practice, (28 practices; 47 clinicians)	213	Intent-to-treat (6 weeks, 97%; 3 months, 96%)	Yes	Low
Rickles et al., 2005	(14 Community pharmacists; 60 clinicians)	63	Intent-to-treat (3 months, 95%)	Yes	Moderate
Rost et al., 2001	(12 Community practices; 24 clinicians) Mixed (8 systems;	479	Intent-to-treat (6 months, 90%; 24 months, 67%)	Yes	Low
Simon et al., 2000	Staff model HMO (5 practices; 40–50 clinicians)	613	Intent-to-treat (3 months, 97%; 6 months, 96%; 60 months, 73%)	Yes	Low
Simon et al., 2004	Staff model HMO (7 practices)	600	Intent-to-treat (6 months, 89%; 18 months, 85%)	Yes	Low

Study	Organization (practices)	n	Study Quality		
			Analysis (follow-up rate)	Blinding outcomes assessment	Risk of bias
Swindle et al., 2003	Veterans Affairs (2 firms; 23 clinicians; 100 trainees)	268	Completers (3 months, 92%; 6 months, 83%)	NS	High
Unutzer et al., 2002	18 practices; 367 clinicians)	1801	Intent-to-treat (3 months, 90%; 6 months, 87%; 12 months, 83%)	Yes	Low
Waterreus et al., 1994	General practice, London, UK (practices, NS; 25 clinicians)	96	Completers (3 months, NS; 6 months, NS; 12 months, NS)	Yes	Moderate
Wells et al., 2000	Staff model and network-managed care (46 practices; 181 clinicians)	1356	Intent-to-treat (6 months, 85%; 12 months, 83%)	Yes	Low

Appendix B. Peer Review Table--Evidence Synthesis For Determining Key Features of Effective Depression Interventions

Section of Report	Actual Comments	Changes
All	Should be reviewed and perhaps rewritten for clarity. Sentence structure difficult to interpret i.e. page 22 first sentence. I had to read this sentence multiple times before I understood intent. (Correlations for evaluation features showed that……..) Consider more sentences but each much shorter.	Re-written and clarified.
Conclusions	"I also do see much of an emphasis that these results are really related to major depression and not subsyndromal or minor depression."	Added
Executive Summary	Suggest that clearly state at the top of page 6 that the outcomes of interest here are related to depression outcome and not other outcomes (HRQoL or mortality), and that analysis relating to comorbidities was limited to psychiatric comorbidities.	On pages 6 and 7, we now clarify that we did aim to look for medical comorbidities, but the studies did not support it. We also clarify that our impact measure reviewed adherence, satisfaction, and functioning. We provide additional detail on these issues later in the document.
Executive Summary	I found the summary to be a bit difficult to read, especially if I had been unfamiliar with the field. Defining terms such as “collaborative care model”, “chronic illness model”, “Patient self management”, etc. early on would help the reader. They are defined later in the work but only after they have been used many times. Page 6: would suggest a word replacement for one of the “interventions”. It would read better.	More definition is now up front. Page 6 sentence fixed.
Executive Summary	"Research questions didn’t address any comorbid medical conditions, which might be very important in a primary care focus (page 6 Primary and Secondary Research Questions)."	Excellent idea, but studies did not include information on medical comorbidities. We have added this to our aims (we did intend to look at this) and limitations.
Executive Summary	"Page 1 gives no reference for their statement “without implementation of organizational changes in primary care practices less than half of patient found to have major depression complete minimally adequate medications or psychotherapy.” "	We now reference this statement.
Executive Summary	"The paper recommends elements of patient activation despite that they only reviewed studies with a patient directed component. Page 5 states: “We did not review studies that only sought to change primary care clinician behavior, without an additional patient-directed component.” Page 7 concludes “patient self-management support as the single statistically significant intervention characteristic associated with improved depression symptoms and depression resolution”----but it seems they excluded all studies without a patient directed component."	We now clarify with examples that the requirement for a patient directed component was not aimed at patient self-management support, but rather at excluding studies focused only on provider-level interventions such as education or decision support. As shown in our tables, but not made clear enough in the write-up, many studies did not include active self-management support.

Introduction	Objectives and scope are clearly described	
Introduction	Suggest a better term [see page 9] than 'mild' major depression, which sounds like a non sequitur; perhaps major depression of lesser severity, which is less likely to be misinterpreted by the reader as minor depression, which is a less compelling clinical and policy target and not the focus of discussion here.	Changed.
Introduction	Suggest transparency would argue for an explicit identification [e.g., personal communication] of the second study referenced at the top of page 10; I suspect the study in question is an analysis from the PRISMe trial.	It isn't PRISMe; it's an AHRQ review just released and is now referenced (Butler).
Introduction	I would take issue with the wording "to fill in for the required depression treatment support functions" at the bottom of page 12. This describes tasks in a strictly co-located rather than collaborative model, and I would recommend something like "to fill in gaps" (in usual practice) or "to support."	Changed.
Introduction	The goals and objectives of the review are clearly stated. The approach to literature review and data synthesis is of interest and compelling.	
Introduction	"This is a little misleading since severe depression is not often included." Regarding statement of collaborative care and major depression with mild symptoms (bottom of page 9)	Changed.
Introduction	"The term nurse is not always right. Not all care management uses nurses" Regarding Fig 2 (page 12)	Changed.
Introduction	"The range of severity also varies from subsyndromal and up though most exclude complex patients" Addition at top of page 14--variation in studies	Changed.
Introduction	"On page 15, it appears that the authors used the Williams review 2007 as the basis for their paper. It appears that the Williams study focused on 28 studies. How does what the authors did differ from what Williams already did?"	We have added to our explanations on this.
Limitations	"Not mentioned is the extensive selection bias that occurs in these studies in regards to participation. Even the studies that use screening exclude many patients or many to no agree to participate. Therefore we don’t really know how well these models can be applied to the general population in a primary care practice."	Added.
Methods	Methods are clearly described	
Methods	No related studies overlooked	
Methods	Page 18: for the short, medium,and long effect size groups, were those that showed effects in more than one group given greater weight on impact analysis?	See above re more information on the impact variable; yes, the impact variable considered how durable and consistent the improvement was.
Methods	Not aware of any related depression studies that have been overlooked	

Methods	"The way I read the exclusion criteria, it isn't clear why the prisme study wasn't included. Krahn, etal. Otherwise the exclusion needs to say that the control group was always usual care"	Added usual care
Methods	"Of the 1464 studies identified, why did only 28 get analyzed? Do the authors feel that the 28 studies, less than half of which were published in the last 5 years, provide an adequate basis to make generalizations, or are the numbers so small that the take home message should be to just to fund more studies? Of the 28 studies analyzed, what were the controls? The studies demonstrate that depression improves with a primary care clinician, a mental health clinician, a care manager, patient self-involvement, but compared to what? Compared to no care? Compared to just a primary clinician alone? Compared to a mental health clinician? Compared to having both primary care and mental health clinicians and no care manager? Compared to all of the above but no patient involvement?"	The article drop-off seen here is typical of evidence review. See Fig. 1 for exclusion reasons. However, we have included more on this under limitations. Actually, 28 randomized trials on a single type of intervention is a much greater number than is available for many or even most quality improvement interventions. Re controls, we have clarified throughout that the comparison is to usual care.
Methods	With respect to any overlooked studies: "Defer to other subject matter experts"	
Methods	VHA program is primary care-mental health integration [page 15; it is often reversed in common usage but technically mental health-primary care in VA is the placement of medical support in outpatient psychiatric venues]. The Uniform Mental Health Services Package is VHA Handbook 1160.01 for your reference.	Switched and referenced.
Results	No bias in synthesis of the evidence	
Results	I would consider adding a statement to the limitation section, which your statistician feels is appropriate, about the possible relationship between the number of categories for analytic variables, misclassification bias between the categories, and the results presented. In other words, comment on reproducibility/robustness given these categories.	Added.

Results	Are the analyzed interventions unique to PC/MH? Page 17 defines care management as: “abstracted features such as : coordination and communication among providers, patient education, monitoring sxs and adherence to Rx Plans, self management support, and psychological Rx.” All of the above seem components of any good treatment plan. Would they have the same impact and effect on any type of disease management program? Individual intervention features were analyzed in the study. Could combinations of interventional variables and their impact yield additional information? Most/Many of the analyzed studies had additional staff support. Could the same clinical effect have been realized by lowering PC panel size to increase patient contact and intervention time?	Excellent points. We now discuss these points in the results and limitations. We don’t know how these interventions would work for other conditions, but we now point out some unique characteristics of depression that might predict more impact for this condition than others. We also point out why, while they are characteristic of any good treatment plan, they are particularly difficult to achieve for depression in primary care. The effect of lowering panel size and increasing time with primary care would be an excellent thing to study; it might, however, be considerably more costly than adding staff for care management (so I was told by Kaiser upper management when it was proposed during my quality improvement study there!).
Results	No indication of overt bias; much potential for subjective bias in this study during abstraction of variables (eg. last paragraph, page 18)	
Results	Page 24 Qualitative Analysis of Intervention Features vs Outcome Impacts: "These were all commonly used so it is hard to say how critical they really are"	Changed language. Basically, we know the features characterized most of the sample of studies, which overall were effective, and we also know they clustered in the most impactful studies. But we don't know that taking any one of them out would undo the effect. Still can say that to mirror the literature, people should adhere to these features, pending additional information.
Results	.[With respect to following Report statement on page 25: No study excluded anxiety. Six studies (21%) did not mention PTSD, and only 14% of the remaining 22 studies excluded patients based on it. Six studies (21%) did not report on the proportion of minorities enrolled.] "Alcohol is also not mentioned in these but very common in practice. I would say more that we know these anxiety disorders, cognition, alcohol use, etc are common and we don’t know the impact on care management"	Added.
Tables - 1	Significant typo at the bottom of page 27--change "diabetes" to "depression"	Changed.
Tables - 6	Table 6 shows studies that all have the same category of supportive interventions. Many have “yes” responses but then are rated anywhere from 1-4 on impact. The study indicates this was a personal review which had a high level of concurrence in rating. Would it have been better to establish some point values which related to the 1-4 level of ratings rather than enter subjective bias into the review?	Explanations were unclear. We have now included much more information on the impact rating, how it worked, and why it was used in Methods and in Results.