APPENDIX A. SEARCH STRATEGY

Step	Category	Terms	Result
1	**Eligible disorders**	("Serious mental illness") [all fields] OR ("severe mental illness") [all fields] OR schizophrenia [tiab] OR schizophrenia [mesh] OR bipolar disorder [mesh:noexp] OR bipolar disorder [tiab] OR psychotic disorders [mesh:noexp] OR psychotic disorders [tiab] OR schizoaffective disorder* [tiab] OR mania [tiab] OR manic [tiab] OR bipolar affective disorder [tiab] OR *mental disorders [tiab]	790929
2	**Interventions**	Delivery of Health Care, Integrated [Mesh] OR Patient Care Team [Mesh] OR Patient Care Planning [Mesh] OR Disease Management [Mesh] OR Comprehensive Health Care [Mesh:noexp] OR Patient Care Management [Mesh:noexp] OR Primary Health Care [Mesh] OR Internal Medicine [Mesh] OR Family practice [Mesh] OR Geriatrics [Mesh] OR "general practice" [ti] OR ("continuity of care" OR "coordinated care" OR "coordinated program*" OR "team care" OR "team treatment" OR "team assessment" OR "team consultation") OR (collaborat*[ti] AND care [ti]) OR "shared care" [ti] OR (collaborat*[ti] AND manage*[ti])	292051
3	**Study designs**	("pre-post" [tiab] OR "pre test" [tiab] OR "pre-test" [tiab] OR "pretest" [tiab] OR "post test" [tiab] OR "post-test" [tiab]) OR ((before[tiab] AND after [tiab]) OR (before [tiab] AND during [tiab])) OR (quasi-experiment*[tiab] OR quasiexperiment*[tiab] OR quasirandom* [tiab] OR quasi random*[tiab] OR quasicontrol* [tiab] OR quasi control* [tiab]) OR ("time series" [tiab] AND interrupt* [tiab]) OR ("time points" [tiab] AND (multiple[tiab] OR three[tiab] OR four[tiab] OR five[tiab] OR six[tiab] OR seven[tiab] OR eight[tiab] OR nine[tiab] OR ten[tiab] OR month*[tiab] OR hour*[tiab] OR day*[tiab])) OR ("process assessment (health care)" [MeSH Terms] OR program evaluation [mesh]) OR ((clinical [tiab] AND trial [tiab]) OR clinical trials [MeSH Terms] OR clinical trial [Publication Type] OR random*[tiab] OR random allocation [MeSH Terms] OR therapeutic use [MeSH Subheading])	3564636
4	**Combine results**	#1 AND #2 AND #3	
5	**Apply limits**	**LIMITS:** English and Human and Adult	1058

APPENDIX B. STUDY SELECTION FORM

Criteria for inclusion and exclusion of studies

Inclusion criteria:

- Study designs recommended by the Cochrane Effective Practice and Organization of Care Group (does NOT include cross-over or observational):
    - Patient or cluster RCTs
    - Nonrandomized cluster controlled trials: An experimental study in which practices or clinicians are allocated to interventions using nonrandom methods
    - Controlled before-and-after studies: A study in which observations are made before and after the implementation of an intervention, both in a group that receives the intervention and in a control group that does not
    - Interrupted time series designs: A study that uses observations at multiple time points before and after an intervention – an attempt to detect if the intervention has had an effect significantly greater than any underlying trend over time
- Sample population has schizophrenia, schizoaffective disorder and/or bipolar disorder, or meets the definition of SMI based on low functional status and least 25% are diagnosed with schizophrenia, schizoaffective disorder and/or bipolar disorder.
- Sample population age 18 and over
- Outpatient population (from mental health clinics and satellite clinics, not community sites)
- Intervention or "exposure" meets definition for integrated care with the explicitly stated goal of improving general medical outcome(s). At a minimum, integrated care must:
    - Involve system redesign such that care providers are added to directly address or coordinate mental and general medical care. Examples include: adding a general medical provider (PA, APN, MD) to the mental health setting, adding a behavioral health specialist who can address multiple behaviors related to general medical care or a health coach /educator /nurse to coordinate and follow through on general medical care with providers located outside the mental health specialty setting.
    - If system redesign with care providers is not used, there must be at least 3 **of the following elements** designed to provide integrated mental and general medical care (decision support, information systems, self-management support, teams care or enhanced communication).
- Includes results on at least one of the relevant outcomes (KQs 1–3)
- Study duration of at least 3 months
- Must be in a peer-reviewed publication
- English language
- Study conducted in North America, Western Europe, Australia/New Zealand

Exclusion criteria:

- Non-English language publication
- Cross-sectional studies and other observational study designs not specifically listed as "included" study designs
- Studies in which the sample is selected for individuals with substance abuse disorders
- Community, rather than practice-based interventions (i.e., not interested in senior centers, but what can be achieved within existing VA clinics and satellite facilities)
- Interventions designed to affect only one specific outcome or aspect of general medical health (e.g., weight loss or smoking cessation, etc.)
- Interventions that involve only: self-management support, enhanced information systems (e.g. EMR, shared records), decision support (e.g., clinical guidelines, clinical reminders) or enhanced access (e.g., location closer to target population or open access scheduling).

APPENDIX C. CRITERIA USED IN QUALITY ASSESSMENT

General Instructions:

For each risk of bias item, rate as "Yes," "No," or "Unclear." After considering each of the quality items, give the study an overall quality rating of good, fair, or poor.

Detailed Quality Items:

If an item is rated as "No," describe why in the comments column.

1. *Randomization adequate? Was the allocation sequence adequately generated?* **Yes/No/Unclear**
2. *Allocation concealment adequate? Was allocation adequately concealed?* **Yes/No/Unclear**
3. *Incomplete outcome data adequately addressed?* **Yes/No/Unclear**
   Consider Attrition bias: Were there systematic differences between groups in withdrawals from a study or high overall loss to followup? (Even small differences could be important when rates are low.) Were subjects excluded from the analysis – if so, were the exclusions sensible?
4. Subjects Blinded? *Were subjects blind to treatment assignment?* **Yes/No/Unclear**
5. Outcome assessor blinded? (This may be recorded separately for each critically important outcome.) *Were* Outcome assessors *blind to treatment assignment?* **Yes/No/Unclear**
6. Provider (treating clinician) blinded? *Were providers blind to treatment assignment?* **Yes/No/Unclear**
7. All outcomes reported? *Are reports of the study free of suggestion of selective outcome reporting (systematic differences between reported and unreported findings)?* **Yes/No/Unclear**
8. Intention-to-treat analysis? **Yes** (all eligible patients that were randomized are included in analysis; note- mixed models and survival analyses are in general ITT) **/No/Unclear**
9. *Adequate power for main effects?* **Yes** (if power analysis or sample size calculation given and recruitment met needs or if post-hoc power calculation shows adequate power)**/No** (did not meet projected sample size needs) **/Unclear** (no power or sample size calculation given)
10. *Other Selection bias? Were there methods that could lead to differences or were there systematic differences observed in baseline characteristics and prognostic factors of the groups compared?(e.g., failure of randomization):* **Yes/No/Unclear**
11. Comparable groups maintained? (Includes crossovers, adherence, and contamination.) Consider issues of crossover (e.g., from one intervention to another), adherence (major differences in adherence to the interventions being compared), contamination (e.g., some members of control group get intervention) Yes/No/Unclear
12. *Lack of Performance bias? Were there no important systematic differences in the care that was provided, other than the intervention of interest?* **Yes/No/Unclear**
13. *Lack of Measurement bias? Were the measures used reliable and valid – and therefore, "yes" no important measurement bias?* **Yes/No/Unclear**
14. *Absence of Detection bias? Were there systematic differences between groups in how outcomes are determined? If no systematic differences answer "yes" – no important detection bias.* **Yes/No/Unclear**
15. *Was there the absence of potential important conflict of interest? The focus here is financial conflict of interest. Therefore if no financial conflict of interest (e.g. funded by government or foundation and authors do not have financial relationships with drug/device manufacturer), then answer "yes."* **Yes/No/Unclear**

Overall rating

Please assign each study an overall quality rating of "Good," "Fair," or "Poor" based on the following definitions:

A "**Good**" study has the least bias, and results are considered valid. A good study has a clear description of the population, setting, interventions, and comparison groups; uses a valid approach to allocate patients to alternative treatments; has a low dropout rate; and uses appropriate means to prevent bias, measure outcomes, and analyze and report results.

A "**Fair**" study is susceptible to some bias but probably not enough to invalidate the results. The study may be missing information, making it difficult to assess limitations and potential problems. As the fair-quality category is broad, studies with this rating vary in their strengths and weaknesses. The results of some fair-quality studies are possibly valid, while others are probably valid.

A "**Poor**" rating indicates significant bias that may invalidate the results. These studies have serious errors in design, analysis, or reporting; have large amounts of missing information; or have discrepancies in reporting. The results of a poor-quality study are at least as likely to reflect flaws in the study design as to indicate true differences between the compared interventions.

Table 11. Quality assessment for the four RCTs

Quality item	Druss et al., 2001	Bauer et al., 2006 and Kilbourne et al., 2009	Kilbourne et al., 2008	Druss et al., 2010
1. Randomization adequate?	Yes	Yes	Yes	Yes
2. Allocation concealment adequate?	Unclear	Unclear	Yes	Yes
3. Incomplete outcome date adequately addressed?	Yes	Yes	Yes	Yes
4. Subject blinded?	No	No	Yes	No
5. Outcome assessor blinded?	Unclear	Unclear	Yes	Yes
6. Provider blinded?	No	No	No	No
7. All outcomes reported?	Yes	Yes	Yes	Yes
8. Intention-to-treat analysis?	Yes	Yes	Yes	Yes
9. Adequate power for main effects?	Unclear	Yes	Unclear	Unclear
10. Other selection bias?	Yes	Yes	Yes	Yes
11. Comparable groups maintained?	Yes	Yes	Yes	Yes
12. Lack of performance bias?	Yes	Yes	Yes	Yes
13. Lack of measurement bias?	Yes	Yes	Yes	Yes
14. Absence of detection bias?	Yes	Yes	Yes	Yes
15. Was there the absence of potential important conflict of interest?	Yes	Yes	Yes	Yes

APPENDIX D. PEER REVIEW COMMENTS/AUTHOR RESPONSES

Reviewer	Comment	Response
Question 1: Are the objectives, scope, and methods for this review clearly described?		
1	Yes	Acknowledged
2	Yes. The report is very clear.	Thank you.
3	Yes	Acknowledged
4	Yes	Acknowledged
5	Yes	Acknowledged
Question 2: Is there any indication of bias in our synthesis of the evidence?		
1	No	Acknowledged
2	No	Acknowledged
3	No	Acknowledged
4	No. It is interesting that three out of the four RCTs were VA studies. While I don't think this indicates bias, I think it does reflect the high quality research being conducted in VA and the cutting-edge nature of what VA does. I'm not sure I agree with the statement that this might impact the applicability of the findings to a non-VA setting, although I appreciate the authors' sensitivity to this issue. The fact that two of the four studies included people with bipolar disorder exclusively is a significant limitation, as was pointed out in the review but gives clear direction for future research and emphasizes the need for VA to use a clear definition for "serious mental illness."	Thank you.
5	No	Acknowledged
Question 3: Are there any studies of interest to the VA that we have overlooked?		
1	Not sure: check this one (I will attach pdf to response email): Miller AL, Crismon ML, Rush AJ, et al. The Texas Medication Algorithm project: Clinical results for schizophrenia. Schizophr Bull. 2004;30(3):627-647.	Thank you for the suggestion, but the Miller study would not have met our inclusion criteria for two reasons: (1) The intervention did not meet our definition for integrated care. (2) The purpose was to improve symptoms of schizophrenia, not medical outcomes. The only nonpsychological measure is the SF-12.
2	No. Not to my knowledge	Acknowledged

Reviewer	Comment	Response
3	Yes. Zappe C, Danton W. Integrated mental health and primary care: a model of coordinated services. Federal Practitioner, 2004. June: 74-81. McGuire J, Gelberg L, Blue-Howells J, Rosenheck RA. Access to primary care for homeless veterans with serious mental illness or substance abuse: a follow-up evaluation of co-located primary care and homeless social services. Adm Policy Ment Health, 2009. 36(4): 255-64. Note: neither meets criteria (first for design, 2nd for proportion of participants with SMI) but might be listed under those reports reviewed but not included	Thank you for the suggestions. The Zappe study did not come up in our literature search because it is not an RCT or other included study design. The McGuire study was found in our literature review but was excluded at the abstract level for population not of interest due to substance abuse. Per systematic review standard protocol, it was not included in the table of excluded studies because it was not reviewed at the full-text level.
4	No	Acknowledged
5	Yes. This is likely, given the current interest in PACT and special populations. The review methodology clearly disqualified QI studies in favor of RCTs. While this has scientific merit, it risks overlooking important and usually unfunded pre/post studies. It is, of course, difficult to get Qi projects published, so the search for successful interventions would be difficult and at variance from the usual processes and definitions of "evidence synthesis."	We agree that valuable information may be contained in quality improvement evaluations of interventions, many of which do not get published. For the purposes of this review, we used the study criteria recommended in the Cochrane Effective Practice and Organization of Care Search. The established criteria for the evidence synthesis included study designs in addition to RCTs; however, we did not identify any non-RCT studies that met the established criteria.
Question 4: Please write additional suggestions or comments below. If applicable, please indicate the page and line numbers from the draft report.		
1	1. Overall very nice job—hard digging for a few nuggets. Important that you point this out to the field. I particularly like that you have policy/funding directives at the end that are fairly specific, not just "more research is needed." I also like that you list the excluded studies. It allowed me to cross-check our own review quickly and see if there was anything we got that you didn't. It's also great that you list the clinicaltrials.gov list of trials in progress so we can watch for "coming attractions."	Thank you.
	2. Related to this, you have on p 13 a separate section on "Rating the Body of Evidence" but I don't see that as an integral and major part of your Recommendations on p 36. I may have missed it, but this may be because it needs further highlighting.	Text has been added to the Summary and Discussion section about the rating of evidence.
	3. You may want to consider adopting the PRISMA reporting system for your ESPs: Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. J Clin Epidemiol. 2009;62(10):e1-34; you get the info mostly there but PRISMA is becoming the standard (eg, by JAMA)	Acknowledged

Reviewer	Comment	Response
1 (cont.)	4. It seems that a very key and important issue is that "None of the four trials provided information on general medical outcomes" (p 29 para 3). It seems this should be the #1 focus for future research. I'm not so concerned that a wide variety of care models haven't been tested—it's a good thing that we can build off of one so strongly supported as the CCM. But at this point we really have no idea whether we can make a dent in the deficits that motivated this review in the first place: premature mortality and poor medical outcome.	We agree and have addressed this point in the first paragraph of the Recommendations for Future Research section.
	5. There is a mis-statement on p 29 in para 4 (that is not consistently made in the document but should be corrected here: "Three of the four studies (54-56, 58, 60) evaluated interventions implemented at only one site." References 55, 56, and 58 refer to an 11-site, 3-year RCT.	The reviewer is correct. We simply referenced the wrong study as being the third of the three studies conducted at one site. The two Druss studies are correct. We have replaced the references with Kilbourne 2008.
	6. The exclusion of PTSD as an SMI seems to me to be influenced by programmatic/policy perspectives rather than clinical. To wit: The VA counts as SMI bipolar spectrum and schizophrnenia spectrum disorders, but not PTSD; the latter has its own tracking system, clinical programs, and champions who by and large differ from those committed to SMI. However, clinically, PTSD is also typically treated in the specialty MH sector, and medication such as second generation antipsychotics which can worsen medical health are used widely. Thus PTSD is characterized by both fragmentation of care and elevated iatrogenic medical risk. Additionally, more Veterans with PTSD are treated annually by VHA (~300K) than with bipolar disorder (~100K) or schizophrenia (~90K) at last count. On the other hand, most published data on "SMI" focus on the narrower definition that you adopt. In either event, if you go back and look for PTSD it's not likely you'll find any studies, although this one may make it: Zatzick D, Roy-Byrne P, Russo J, et al. A randomized effectiveness trial of stepped collaborative care for acutely injured trauma survivors. Arch Gen Psychiatry. 2004;61(5):498.	We agree that PTSD is an extremely important diagnostic entity, particularly in the VA, and acknowledge the similarities with the disorders emphasized in this review. Diagnostic inclusion criteria were informed by the views of our identified stakeholders. Two of the included studies did have subjects with PTSD. The study cited in the comment is on subjects with PTSD due to physical assault. Some are hospitalized. A good proportion have substance abuse. There are no medical outcomes. While it is an important study, it does not meet the criteria established for this review.
	7. It is likely worth noting that CCMs have begun to enter the clinical practice guideline literature as a fundamental approach to care for SMI—specifically bipolar disorder. Here are two instances, and I believe the draft of the American Psychiatric Association Guideline for bipolar disorder will cite the model as well: VA-DoD: Department of Veterans Affairs & Department of Defense. Clinical practice guideline for management of bipolar disorder in adults, version 2.0. Department of Veterans Affairs Office of Quality and Performance & US Army MEDCOM Quality Management Division. 2009. CANMAT: Yatham LN, Kennedy SH, O'Donovan C, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) guidelines for the management of patients with bipolar disorder: Update 2007. Bipolar Disord. 2006;8(6):721-739	We have added mention of these guidelines in the summary and discussion of KQ 4.

Reviewer	Comment	Response
2	No specific suggestions/comments.	Acknowledged
3	The report is surprisingly lengthy given the paucity of literature on the topic. It is, however, comprehensive, and detailed on the information that is available. A less restrictive selection strategy may have allowed for more comment (though perhaps in a less definitive manner) on "real world" application of these care models (see 5th item on Table 8, page 29).	Acknowledged
4	While I understand the definition of "integrated care" used in the review, I would not have included "health coach/educator," especially when it comes to providing primary care to individuals with SMI. From my experience, such providers are unprepared to work with individuals with highly complex mental health needs. In addition, "coordinate and follow through" services are qualitatively different from directly providing primary care services in specialty mental health or mental health services in primary care. Including a "health coach/educator" in the definition unnecessarily complicates the issue.	We agree that working with individuals with SMI requires a complex set of skills. The type of health professionals noted in this comment would not provide primary care services to the exclusion of other team members in the models included. We defined the features of integrated care using the chronic care model and medical home model as guides. Therefore, our inclusion criterion for the intervention was, "… meets definition for integrated care with the explicitly stated goal of improving general medical outcome(s). At a minimum, integrated care must: (1) Involve system redesign such that care providers are added to directly address or coordinate mental and general medical care. Examples include: adding a general medical provider (PA, APN, MD) to the mental health setting, adding a behavioral health specialist who can address multiple behaviors related to general medical care … or a health coach/educator/nurse to coordinate and follow through on general medical care with providers located outside the mental health specialty setting …"
4	The future directions section seems very much on target. Just as studying the addition of primary care services to Assertive Community Treatment programs could yield interesting findings, so too could the addition of primary care services to Psychosocial Rehabilitation and Recovery Centers, which is another piece of the continuum of care from inpatient to outpatient care. This review is very timely, as OMHS is working to define the interaction of primary care and specialty mental health.	Thank you. We have added mention of PRRCs to the Recommendations for Future Research section.

Reviewer	Comment	Response
5	KQ 4 is the most important question for a developing field of knowledge. RCT evidence to date is so limited that it is difficult to make any conclusions other than "we need to know more." These studies have demonstrated that directly addressing general medical needs in the mental health setting is associated with better processes of care that should lead to better outcomes (we also know that improved health outcomes may not be apparent in the timeline associated with RCTs). Given the dearth of information, the literature to date can best be used to confidently state that doing something is better than doing nothing. Given the variability (Kilbourne, Post et al. 2008) in defining "Serious Mental Illness," we need agreement about a research definition of the term that can be applied across future studies.	We agree with the points in this comment.
	The review notes, but does not emphasize the apparent lack of focus on providing PCMH services within existing/developing PCMH (PACT) programs. Does this imply an assumption that it cannot be done? Are there specific interventions that can assure that patients with SMI can receive care in a VA PACT? Creating SMI PACTS in VA mental health services may be possible in a research environment but is likely to be financially unsustainable. SCAN/ECHO is a model that suggests that, with the right supports/education/mentoring in place, general medical practices can successfully treat complex populations.	We do not assume that PACT cannot successfully address the needs of individuals with SMI. The studies reviewed had a treatment-as-usual condition that was not consistent with PACT even though 3 of 4 studies did occur in the VA. We have added text to the KQ 1 discussion to reflect uncertainty about whether PACT can work with individuals with SMI.
5	There is also need to explore models useful in CBOCs. Many CBOCs, by virtue of their small size, have developed fully integrated care programs, though without calling them programs. There is likely a wealth of information about what has been helpful, that could help identify interventions that could then be tested in RCTs.	We have added some discussion of CBOCs in the Recommendations for Future Research section.
Question 5. Are there any clinical performance measures, programs, quality improvement measures, patient care services, or conferences that will be directly affected by this report? If so, please provide detail.		
1	The Mental Health QUERI SMI Health Work Group will be very interested in this (Dr. Williams is a member so I'm sure they will be in the loop).	Acknowledged
2	As the report indicates, this evidence synthesis is highly relevant to the Patient Centered Medical Home (PACT) initiative. This report will be immensely useful to the strategic planning of the Mental Health QUERI SMI Health Workgroup.	Thank you.
3	Not directly	Acknowledged
4	No comment	Acknowledged
5	The Primary Care – Mental Health Integration program has had its lens focused almost entirely on provision of MH services in PC. These services have been mostly limited to care of common, relatively straightforward psychological, psychiatric and social problems. This report may be helpful in expanding those horizons. Likewise, the added emphasis on the patient population most likely to negatively impact on any given PACT performance measure will be important. The studies reviewed note very specific target conditions, which are good fodder for local QI initiatives	Acknowledged

Reviewer	Comment	Response
Question 6: Please provide any recommendations on how this report can be revised to more directly address or assist implementation needs.		
1	1. CCMs per se are really not on the radar of OMHS and this report indicates that they should be. 2. This report also highlights the need for better tracking of quality of care processes for SMI Veterans. Specifically, there have been overlapping/colliding efforts across OMHS and OQP to develop performance measures around metabolic monitoring and SMI (with/without antipsychotic use). Amy Kilbourne was leading this nascent effort that I think has, unfortunately, died on the vine. I would hope that the recommendations of this report might reinvigorate this effort.	Acknowledged
2	No revisions are needed.	Thank you.
3	Unfortunately, the literature review suggest benefits but, as laid out clearly in the report, there are many gaps remaining in knowledge on this topic.	Acknowledged
4	No comment	Acknowledged
5	Like the AHRQ funded reports a few years ago, cautioning against "premature orthodoxy" is important.	We have added text to address this in the KQ 4 discussion.
Question 7: Please provide us with contact details of any additional individuals/stakeholders who should be made aware of this report.		
1	Not sure who you are already going to contact. The Usual Suspects probably include MH QUERI and OMHS. Diabetes QUERI and related medically oriented QUERIs also come to mind. OQP (or whatever it's called now) Grant Huang the head of CSP, since you recommend a CSP-level trial The SMI Committee in particular under OMHS Outside VA: NAMI and the Depression & Bipolar Support Alliance	Thank you for the suggestions. We will disseminate the report in these directions.
2	I believe the key stakeholders have already been included in developing this report, including OMHS, HSR&D/QUERI, and Mental Health QUERI. The Primary Care-Mental Health Integration Initiative and PACT leaders should be made aware if they are not already on the list.	Thank you for the suggestions. We will disseminate the report in these directions.
3	No comment	Acknowledged
4	No comment	Acknowledged
5	Jeff Burk, national director of psychosocial rehab and recovery is vital to this area. If not already reviewing, he should be added	Thank you for the suggestion. Dr. Burk was a reviewer of this report. We will make sure he is aware of the final report.

APPENDIX E. EXCLUDED STUDIES

All studies listed below were reviewed in their full-text version and excluded for the reason indicated. An alphabetical reference list follows the table.

Reference	Not SMI	Not outpatient	Not RCT	Not integrated care	No medical outcomes	Not peer-reviewed	Not Westernized culture
Adair et al., 2005 (1039)			X				
Baker et al., 2009 (1055)			X				
Bauer et al., 2001 (1592)				X			
Bauer et al., 2007 (1558)						X	
Byng et al., 2004 (434)					X		
Chafetz et al., 2008 (152)		X					
Chiverton et al., 2007 (185)	X						
Ciompi et al., 1992 (907)			X				
Davies et al., 2008 (1134)			X				
Desai et al., 2002 (16)	X						
Desai et al., 2002 (23)	X						
Dewa et al., 2009 (82)	X						
Dickerson et al., 2003 (11)			X				
Donald et al., 2005 (395)			X				
Drew et al., 2007 (217)			X				
Druss et al., 2010 (21)				X			
Essock et al., 1998 (687)					X		
Essock et al., 1995 (820)					X		
Essock et al., 2006 (319)				X			
Forsberg et al., 2008 (1553)				X			
Harvey et al., 2005 (1200)				X			
Jerrell et al., 1995 (806)					X		
Kahn et al., 2009 (46)	X						
Kalichman et al., 1995 (832)				X			

Reference	Not SMI	Not outpatient	Not RCT	Not integrated care	No medical outcomes	Not peer-reviewed	Not Westernized culture
Katon et al., 1991 (929)	X						
Kemp et al., 2010 (1234)				X			
Know et al., 2006 (1267)							X
Madhusoodanam et al., 2006 (1298)			X				
Malla et al., 1998 (675)				X			
McKibbin et al., 2010 (1324)				X			
Ohlsen et al., 2005 (341)			X				
O'Kearney et al., 2004 (1362)			X				
Pirraglia et al., 2009 (1380)			X				
Poulin et al., 2007 (1383)				X			
Ridgely et al., 1996 (774)			X				
Rivera et al., 2007 (225)				X			
Robson et al., 1984 (960)	X						
Rubin et al., 2005 (376)		X					
Ryan et al., 2007 (207)	X						
Sartorius et al., 1993 (887)	X						
Sata et al., 1999 (614)	X						
Schmidt-Kraepelin et al., 2009 (55)					X		
Sim et al., 2006 (1446)				X			
Simon et al., 2006 (1694)				X			
Snyder et al., 2008 (1693)				X			
Symonds et al., 2007 (1692)	X						
Taborda et al., 2003 (471)						X	
Thompson et al., 2006 (1484)				X			
Welch et al., 2009 (84)			X				
Wright et al., 2006 (308)			X				

LIST OF EXCLUDED STUDIES

Adair CE, McDougall GM, Mitton CR, et al. Continuity of Care and Health Outcomes Among Persons With Severe Mental Illness. *Psychiatr Serv*. 2005;56(9):1061-1069.

Baker A, Richmond R, Castle D, et al. Coronary heart disease risk reduction intervention among overweight smokers with a psychotic disorder: Pilot trial. *Aust N Z J Psychiatry*. 2009;43(2):129-135.

Bauer MS. The collaborative practice model for bipolar disorder: design and implementation in a multi-site randomized controlled trial. *Bipolar Disorders*; 2001:233-44.

Bauer Ms KAM. Outcome and Costs in a Randomized Controlled Effectiveness Trial of a Collaborative Chronic Care Model for Bipolar Disorder. *Journal of Mental Health Policy and Economics*; 2007:S3.

Byng R, Jones R, Leese M, et al. Exploratory cluster randomised controlled trial of shared care development for long-term mental illness. *Br J Gen Pract*. 2004;54(501):259-66.

Chafetz L, White M, Collins-Bride G, et al. Clinical trial of wellness training: health promotion for severely mentally ill adults. *J Nerv Ment Dis*. 2008;196(6):475-83.

Chiverton P, Lindley P, Tortoretti DM, et al. Well balanced: 8 steps to wellness for adults with mental illness and diabetes. *J Psychosoc Nurs Ment Health Serv*. 2007;45(11):46-55.

Ciompi L, Dauwalder HP, Maier C, et al. The pilot project 'Soteria Berne'. Clinical experiences and results. *Br J Psychiatry Suppl*. 1992(18):145-53.

Davies LM, Barnes TRE, Jones PB, et al. A randomized controlled trial of the cost-utility of second-generation antipsychotics in people with psychosis and eligible for clozapine. *Value in Health*. 2008;11(4):549-562.

Desai MM, Rosenheck RA, Druss BG, et al. Mental disorders and quality of care among postacute myocardial infarction outpatients. *J Nerv Ment Dis*. 2002;190(1):51-3.

Desai MM, Rosenheck RA, Druss BG, et al. Receipt of nutrition and exercise counseling among medical outpatients with psychiatric and substance use disorders. *J Gen Intern Med*. 2002;17(7):556-60.

Dewa CS, Hoch JS, Carmen G, et al. Cost, effectiveness, and cost-effectiveness of a collaborative mental health care program for people receiving short-term disability benefits for psychiatric disorders. *Can J Psychiatry*. 2009;54(6):379-88.

Dickerson FB, McNary SW, Brown CH, et al. Somatic healthcare utilization among adults with serious mental illness who are receiving community psychiatric services. *Med Care*. 2003;41(4):560-70.

Donald M, Dower J, Kavanagh D. Integrated versus non-integrated management and care for clients with co-occurring mental health and substance use disorders: a qualitative systematic review of randomised controlled trials. *Soc Sci Med*. 2005;60(6):1371-83.

Drew L, Delacy F. Improving the general health of persons with psychosis. *Australas Psychiatry*. 2007;15(4):320-3.

Druss BG, Zhao L, von Esenwein SA, et al. The Health and Recovery Peer (HARP) Program: A peer-led intervention to improve medical self-management for persons with serious mental illness. *Schizophr Res*. 2010.

Essock SM, Frisman LK, Kontos NJ. Cost-effectiveness of assertive community treatment teams. *Am J Orthopsychiatry*. 1998;68(2):179-90.

Essock SM, Kontos N. Implementing assertive community treatment teams. *Psychiatr Serv*. 1995;46(7):679-83.

Essock SM, Mueser KT, Drake RE, et al. Comparison of ACT and standard case management for delivering integrated treatment for co-occurring disorders. *Psychiatr Serv*. 2006;57(2):185-96.

Forsberg KA, Björkman T, Sandman PO, et al. Physical health--a cluster randomized controlled lifestyle intervention among persons with a psychiatric disability and their staff. *Nordic Journal of Psychiatry*; 2008:486-95.

Harvey SB, Newton A, Moye GA. Physical health monitoring in schizophrenia: The use of an invitational letter in a primary care setting. *Primary Care & Community Psychiatry*. 2005;10(2):71-74.

Jerrell JM, Ridgely MS. Gender differences in the assessment of specialized treatments for substance abuse among people with severe mental illness. *J Psychoactive Drugs*. 1995;27(4):347-55.

Kahn LS, Fox CH, Carrington J, et al. Telephonic nurse case management for patients with diabetes and mental illnesses: a qualitative perspective. *Chronic Illn*. 2009;5(4):257-67.

Kalichman SC, Sikkema KJ, Kelly JA, et al. Use of a brief behavioral skills intervention to prevent HIV infection among chronic mentally ill adults. *Psychiatr Serv*. 1995;46(3):275-80.

Katon WJ. The development of a randomized trial of consultation-liaison psychiatry trial in distressed high utilizers of primary care. *Psychiatr Med*. 1991;9(4):577-91.

Kemp DE, Gao K, Chan PK, et al. Medical comorbidity in bipolar disorder: Relationship between illnesses of the endocrine metabolic system and treatment outcome. *Bipolar Disorders*. 2010;12(4):404-413.

Kwon JS, Choi J-S, Bahk W-M, et al. Weight Management Program for Treatment-Emergent Weight Gain in Olanzapine-Treated Patients With Schizophrenia or Schizoaffective Disorder: A 12-Week Randomized Controlled Clinical Trial. *J Clin Psychiatry*. 2006;67(4):547-553.

Madhusoodanan S, Moise D, Sajatovic M. Schizoaffective disorder in the elderly. In: French DP, ed. *Schizophrenic psychology: New research.* Hauppauge, NY US: Nova Science Publishers; 2006:57-82.

Malla AK, Norman RM, McLean TS, et al. An integrated medical and psychosocial treatment program for psychotic disorders: patient characteristics and outcome. *Can J Psychiatry*. 1998;43(7):698-705.

McKibbin CL, Golshan S, Griver K, et al. A healthy lifestyle intervention for middle-aged and older schizophrenia patients with diabetes mellitus: A 6-month follow-up analysis. *Schizophr Res*. 2010;121(1-3):203-206.

Ohlsen RI, Peacock G, Smith S. Developing a service to monitor and improve physical health in people with serious mental illness. *J Psychiatr Ment Health Nurs*. 2005;12(5):614-9.

O'Kearney R, Garland G, Welch M, et al. Factors predicting program fidelity and delivery of an early intervention program for first episode psychosis in rural Australia. *AeJAMH (Australian e-Journal for the Advancement of Mental Health)*. 2004;3(2).

Pirraglia PA, Biswas K, Kilbourne AM, et al. A prospective study of the impact of comorbid medical disease on bipolar disorder outcomes. *J Affect Disord*. 2009;115(3):355-359.

Poulin M-J, Chaput J-P, Simard V, et al. Management of antipsychotic-induced weight gain: Prospective naturalistic study of the effectiveness of a supervised exercise programme. *Aust N Z J Psychiatry*. 2007;41(12):980-989.

Ridgely MS, Morrissey JP, Paulson RI, et al. Characteristics and activities of case managers in the RWJ Foundation Program on chronic mental illness. *Psychiatr Serv*. 1996;47(7):737-43.

Rivera JJ, Sullivan AM, Valenti SS. Adding consumer-providers to intensive case management: does it improve outcome? *Psychiatr Serv*. 2007;58(6):802-9.

Robson MH, France R, Bland M. Clinical psychologist in primary care: controlled clinical and economic evaluation. *Br Med J (Clin Res Ed)*. 1984;288(6433):1805-8.

Rubin AS, Littenberg B, Ross R, et al. Effects on processes and costs of care associated with the addition of an internist to an inpatient psychiatry team. *Psychiatr Serv*. 2005;56(4):463-7.

Ryan T, Hatfield B, Sharma I. Outcomes of referrals over a six-month period to a mental health gateway team. *J Psychiatr Ment Health Nurs*. 2007;14(6):527-34.

Sartorius N, Ustun TB, Costa e Silva JA, et al. An international study of psychological problems in primary care. Preliminary report from the World Health Organization Collaborative Project on 'Psychological Problems in General Health Care'. *Arch Gen Psychiatry*. 1993;50(10):819-24.

Sata M, Yoshitake K, Utsunomiya H, et al. Factors affecting disability in patients attending the internal medicine departments of general hospitals. *Psychiatry Clin Neurosci*. 1999;53(6):611-20.

Schmidt-Kraepelin C, Janssen B, Gaebel W. Prevention of rehospitalization in schizophrenia: results of an integrated care project in Germany. *European Archives of Psychiatry and Clinical Neuroscience*. 2009;259 Suppl 2:S205-12.

Sim K, Chan YH, Chua TH, et al. Physical comorbidity, insight, quality of life and global functioning in first episode schizophrenia: A 24-month, longitudinal outcome study. *Schizophr Res*. 2006;88(1-3):82-89.

Simon GE, Ludman EJ, Bauer MS, et al. Long-term effectiveness and cost of a systematic care program for bipolar disorder. *Arch Gen Psychiatry*. 2006;63(5):500-8.

Snyder K, Dobscha SK, Ganzini L, et al. Clinical outcomes of integrated psychiatric and general medical care. *Community Ment Health J*. 2008;44(3):147-54.

Symonds D, Parker R. The Top End Mental Health Services General Practice Clinic: an initiative for patients with serious mental illness. *Australas Psychiatry*. 2007;15(1):58-61.

Taborda JG, Bordignon S, Bertolote JM, et al. Heart transplantation and schizophrenia. *Psychosomatics*. 2003;44(3):264-5.

Thompson WK, Kupfer DJ, Fagiolini A, et al. Prevalence and Clinical Correlates of Medical Comorbidities in Patients With Bipolar I Disorder: Analysis of Acute-Phase Data From a Randomized Controlled Trial. *J Clin Psychiatry*. 2006;67(5):783-788.

Welch K. GPs have a central role in managing schizophrenia. *Practitioner*. 2009;253(1718):19-23.

Wright CA, Osborn DP, Nazareth I, et al. Prevention of coronary heart disease in people with severe mental illnesses: a qualitative study of patient and professionals' preferences for care. *BMC Psychiatry*. 2006;6:16.

APPENDIX F. GLOSSARY

Abstract screening

The stage in a systematic review during which titles and abstracts of articles identified in the literature search are screened for inclusion or exclusion based on established criteria. Articles that pass the abstract screening stage are promoted to the full-text review stage.

ClinicalTrials.gov

A registry and results database of federally and privately supported clinical trials conducted in the United States and around the world. ClinicalTrials.gov provides information about a trial's purpose, location, participant characteristics, among other details.

Cochrane Database of Systematic Reviews

A bibliographic database of peer-reviewed systematic reviews and protocols prepared by the Cochrane Review Groups in The Cochrane Collaboration.

Companion article

A companion article is a publication from a trial that is not the paper containing the main results of that trial. It may be a methods paper, a report of subgroup analyses, a report of combined analyses, or other auxiliary topic that adds information to the interpretation of the main paper.

Confidence interval (CI)

The range in which a particular result (such as a laboratory test) is likely to occur for everyone who has a disease. "Likely" usually means 95 percent of the time. Clinical research studies are conducted on only a certain number of people with a disease rather than all the people who have the disease. The study's results are true for the people who were in the study but not necessarily for everyone who has the disease. The confidence interval is a statistical estimate of how much the study findings would vary if other different people participated in the study. A confidence interval is defined by two numbers, one lower than the result found in the study and the other higher than the study's result. The size of the confidence interval is the difference between these two numbers.

Data abstraction

The stage of a systematic review that involves a pair of trained researchers extracting reported findings specific to the research questions from the full-text articles that met the established inclusion criteria. These data form the basis of the evidence synthesis.

Exclusion criteria

The criteria, or standards, set out before a study or review. Exclusion criteria are used to determine whether a person should participate in a research study or whether an individual study should be excluded in a systematic review. Exclusion criteria may include age, previous treatments, and other medical conditions.

Full-text review

The stage of a systematic review in which a pair of trained researches evaluates the full-text of study articles for potential inclusion in the review.

GRADE

Grading of Recommendations Assessment, Development and Evaluation (GRADE), a system of assessing the quality of medical evidence and evaluating the strength of recommendations based on the evidence.

Inclusion criteria

The criteria, or standards, set out before the systematic review. Inclusion criteria are used to determine whether an individual study can be included in a systematic review. Inclusion criteria may include population, study design, gender, age, type of disease being treated, previous treatments, and other medical conditions.

Nonrandomized study

Any quantitative study estimating the effectiveness of an intervention (harm or benefit) that does not use randomization to allocate units to comparison groups (including studies where "allocation" occurs in the course of usual treatment decisions or peoples' choices; i.e., studies usually called "observational"). There are many possible types of nonrandomized intervention studies, including cohort studies, case-control studies, controlled before-and-after studies, interrupted-time-series studies, and controlled trials that do not use appropriate randomization strategies (sometimes called quasi-randomised studies).

Observational study

A study in which the investigators do not seek to intervene but simply observe the course of events. Changes or differences in one characteristic (e.g., whether or not people received the intervention of interest) are studied in relation to changes or differences in other characteristics (e.g., whether or not they died), without action by the investigator. Observational studies provide weaker empirical evidence than do experimental studies because of the potential for large confounding biases to be present when there is an unknown association between a factor and an outcome.

PsycINFO®

An abstracting and indexing database of peer-reviewed literature in the behavioral sciences and mental health.

Publication bias

The tendency of researchers to publish experimental findings that have a positive result, while not publishing the findings when the results are negative or inconclusive. The effect of publication bias is that published studies may be misleading. When information that differs from that of the published study is not known, people are able to draw conclusions using only information from the published studies.

PubMed®

A database of citations for biomedical literature from MEDLINE®, life science journals, and online books in the fields of medicine, nursing, dentistry, veterinary medicine, the health care system, and preclinical sciences.

Quasi-experimental study

Often described as a nonrandomized, pre-post intervention study. A study based on a true experimental design meets two criteria: manipulation of a variable factor between two or more groups and random assignment of participants to those groups. A quasi-experimental study uses the first criterion, but participants are not randomly assigned to groups. This means a researcher cannot draw conclusions about cause and effect. Quasi-experimental study designs are frequently used when it is not logistically feasible or ethical to conduct a randomized controlled trial.

Randomized controlled trial

A prospective, analytical, experimental study using primary data generated in the clinical environment. Individuals similar at the beginning of the trial are randomly allocated to two or more treatment groups and the outcomes the groups are compared after sufficient followup time. Properly executed, the RCT is the strongest evidence of the clinical efficacy of preventive and therapeutic procedures in the clinical setting.

Risk

A way of expressing the chance that something will happen. It is a measure of the association between exposure to something and what happens (the outcome). Risk is the same as probability, but it usually is used to describe the probability of an adverse event. It is the rate of events (such as breast cancer) in the total population of people who could have the event (such as women of a certain age).

Serious mental illness (SMI)

Defined in this report according to the definition stipulated in Public Law (P.L.) 102–321; that is, a diagnosable mental, behavioral, or emotional disorder, at some time during the past year, that met the criteria in the *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (DSM-IV) (American Psychiatric Association, 1994) and resulted in functional impairment that substantially interfered with or limited one or more major life activities.

Statistical significance

A mathematical technique to measure whether the results of a study are likely to be true. Statistical significance is calculated as the probability that an effect observed in a research study is occurring because of chance. Statistical significance is usually expressed as a P-value. The smaller the P-value, the less likely it is that the results are due to chance (and more likely that the results are true). Researchers generally believe the results are probably true if the statistical significance is a P-value less than 0.05 ($p<.05$).

Strength of evidence (SOE)

A measure of how confident reviewers are about decisions that may be made based on a body of evidence. SOE is evaluated using one of four grades: (1) *High* confidence that the evidence reflects the true effect; further research is very unlikely to change reviewer confidence in the estimate of effect; (2) *moderate* confidence that the evidence reflects the true effect; further research may change the confidence in the estimate of effect and may change the estimate; (3) *low* confidence that the evidence reflects the true effect; further research is likely to change the confidence in the estimate of effect and is likely to change the estimate; and (4) *insufficient*; the evidence either is unavailable or does not permit a conclusion.

Systematic review

A summary of the clinical literature. A systematic review is a critical assessment and evaluation of all research studies that address a particular clinical issue. The researchers use an organized method of locating, assembling, and evaluating a body of literature on a particular topic using a set of specific criteria. A systematic review typically includes a description of the findings of the collection of research studies. The systematic review may also include a quantitative pooling of data, called a meta-analysis.

Time-series study

A quasi-experimental research design in which periodic measurements are made on a defined group of individuals both before and after implementation of an intervention. Time series studies are often conducted for the purpose of determining the intervention or treatment effect.