APPENDIX A. SEARCH STRATEGIES

Search strategy for RCTs and observational studies (PubMed, May 2012)

Step	Category	Terms	Results
1	Fitness center terms	"Health Promotion/economics"[Mesh]) OR "Insurance Benefits"[Mesh]) OR "Insurance Claim Review"[Mesh]) OR "Insurance, Health"[Mesh]	**116141**
2		"Fitness Centers"[Mesh]	**281**
3		(fitness[tiab] OR health[tiab] OR exercise[tiab] OR recreation[tiab] OR recreational[tiab] OR sports[tiab] OR aquatic[tiab]) AND (club[tiab] OR clubs[tiab] OR membership[tiab] OR memberships[tiab] OR center[tiab] OR centers[tiab] OR program[tiab] OR programs[tiab])	**151793**
4		gym[tiab] OR gyms[tiab] OR ymca[tiab] OR "community based"[tiab]	**29217**
5		**#1 AND (#2 OR #3 OR #4)**	**8751**
6	Study design terms	(randomized controlled trial[pt] OR controlled clinical trial[pt] OR randomized[tiab] OR randomised[tiab] OR randomization[tiab] OR randomisation[tiab] OR placebo[tiab] OR drug therapy[sh] OR randomly[tiab] OR trial[tiab] OR groups[tiab] OR Clinical trial[pt] OR "clinical trial"[tw] OR "clinical trials"[tw] OR "evaluation studies"[Publication Type] OR "evaluation studies as topic"[MeSH Terms] OR "evaluation study"[tw] OR evaluation studies[tw] OR "intervention studies"[MeSH Terms] OR "intervention study"[tw] OR "intervention studies"[tw] OR "case-control studies"[MeSH Terms] OR "case-control"[tw] OR "cohort studies"[MeSH Terms] OR cohort[tw] OR "longitudinal studies"[MeSH Terms] OR "longitudinal"[tw] OR longitudinally[tw] OR "prospective"[tw] OR prospectively[tw] OR "retrospective studies"[MeSH Terms] OR "retrospective"[tw] OR "follow up"[tw] OR "comparative study"[Publication Type] OR "comparative study"[tw] OR systematic[subset] OR "meta-analysis"[Publication Type] OR "meta-analysis as topic"[MeSH Terms] OR "meta-analysis"[tw] OR "meta-analyses"[tw]) NOT (Editorial[ptyp] OR Letter[ptyp] OR Case Reports[ptyp] OR Comment[ptyp]) NOT (animals[mh] NOT humans[mh])	**4383508**
7		**#5 AND #6**	**3601**

APPENDIX B. EXCLUDED STUDIES

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

Excluded studies with reasons

Reference	Not full publication, peer-reviewed, or primary data	Not English language	Not geographic location of interest	Not intervention of interest	Not study design of interest	Not outcome of interest
Abildso, 2010					X	
Ackermann, 2003				X		
Ackermann, 2008				X		
Anderson, 2001	X					
Anonymous, 2008	X					
Arlton, 1986	X					
Atherly, 2011				X		
Bartlett-Prescott, 2005				X		
Bertera, 1990	X					
Breuleux, 1993				X		
Burnes, 1995	X					
Compton, 2006				X		
Cooper, 2012						X
Cox, 1981				X		
Deitz, 2005				X		
Fielding, 1982	X					
Foote, 2006	X					
Gillman, 2001				X		
Goetzel, 2001	X					
Goetzel, 1998				X		
Haltiwanger, 2007	X					
Hochart, 2011				X		
Kulesher, 2005	X					
Lambert, 2009			X			
Mayer, 2010				X		
Nguyen, 2007				X		
Orme-Johnson, 1997				X		
Ozminkowski, 2001	X					
Ozminkowski, 2006				X		
Ozminkowski, 2002				X		
Patel, 2011			X			
Patel, 2010			X			
Pelletier, 2005				X		
Pronk, 2002				X		
Serxner, 2001				X		
Shephard, 1982				X		
Shephard, 1983				X		
Shephard, 1982				X		
Spencer, 1996				X		
Spilman, 1986				X		
Stevens, 1998				X		
Terry, 1991				X		
Wylie-Rosett, 2001				X		

Note: At the request of a peer reviewer, we reconsidered one reference[21] that was initially excluded at the title/abstract level; however, we retained our original conclusion that the reference could not be included based on our inclusion/exclusion criteria.

LIST OF EXCLUDED STUDIES

Agency for Healthcare Research and Quality. Group Abildso CG, Zizzi SJ, Reger-Nash B. Evaluating an insurance-sponsored weight management program with the RE-AIM Model, West Virginia, 2004-2008. *Prev Chronic Dis*. 2010;7(3):A46.

Ackermann RT, Cheadle A, Sandhu N, et al. Community exercise program use and changes in healthcare costs for older adults. *Am J Prev Med*. 2003;25(3):232-7.

Ackermann RT, Williams B, Nguyen HQ, et al. Healthcare cost differences with participation in a community-based group physical activity benefit for medicare managed care health plan members. *J Am Geriatr Soc*. 2008;56(8):1459-65.

Anderson DR, Serxner SA, Gold DB. Conceptual framework, critical questions, and practical challenges in conducting research on the financial impact of worksite health promotion. *Am J Health Promot*. 2001;15(5):281-8.

Anonymous. Active seniors decrease health costs: Silver Sneakers study shows significant benefits in year two. *Dis Manag Advis*. 2008;14(5):7-9, 1.

Arlton D. A paying health promotion clinic: combining client services and student learning. *J Allied Health*. 1986;15(1):3-10.

Atherly A, Thorpe KE. Analysis of the treatment effect of Healthways' Medicare Health Support Phase 1 Pilot on Medicare costs. *Popul Health Manag*. 2011;14 Suppl 1:S23-8.

Bartlett-Prescott JD, Klesges LM, Kritchevsky SB. Health promotion referrals in an urban clinic: removing financial barriers influences physician but not patient behavior. *Am J Health Promot*. 2005;19(5):376-82.

Bertera RL. The effects of workplace health promotion on absenteeism and employment costs in a large industrial population. *Am J Public Health*. 1990;80(9):1101-5.

Breuleux C, Heck SK, Hollenback J, et al. Preliminary comparison of medical care costs between fitness center members and nonmembers. *Am J Health Promot*. 1993;7(6):405-7.

Burnes H. Personal health improvement program. *HMO Pract*. 1995;9(2):59-60.

Compton MT, Weiss PS, Phillips VL, et al. Determinants of health plan membership among patients in routine U.S. psychiatric practice. *Community Ment Health J*. 2006;42(2):197-204.

Cooper AL, Trivedi AN. Fitness memberships and favorable selection in Medicare Advantage plans. *N Engl J Med*. 2012;366(2):150-7.

Cox M, Shephard RJ, Corey P. Influence of an employee fitness programme upon fitness, productivity and absenteeism. *Ergonomics*. 1981;24(10):795-806.

Deitz D, Cook R, Hersch R. Workplace health promotion and utilization of health services: follow-up data findings. *J Behav Health Serv Res*. 2005;32(3):306-19.

Fielding JE. Effectiveness of employee health improvement programs. *J Occup Med*. 1982;24(11):907-16.

Foote SM. Medicare health support: reinventing chronic care. *Am Heart Hosp J*. 2006;4(1):39-42.

Gillman MW, Pinto BM, Tennstedt S, et al. Relationships of physical activity with dietary behaviors among adults. *Prev Med*. 2001;32(3):295-301.

Goetzel RZ. The financial impact of health promotion and disease prevention programs--why is it so hard to prove value? *Am J Health Promot*. 2001;15(5):277-80.

Goetzel RZ, Dunn RL, Ozminkowski RJ, et al. Differences between descriptive and multivariate estimates of the impact of Chevron Corporation's Health Quest Program on medical expenditures. *J Occup Environ Med*. 1998;40(6):538-45.

Haltiwanger R. Medicare health support program: better quality of life for chronically ill seniors. *Tenn Med*. 2007;100(7):33.

Hochart C, Lang M. Impact of a comprehensive worksite wellness program on health risk, utilization, and health care costs. *Popul Health Manag*. 2011;14(3):111-6.

Kulesher RR. Medicare-the development of publicly financed health insurance: Medicare's impact on the nation's health care system. *Health Care Manag (Frederick)*. 2005;24(4):320-9.

Lambert EV, da Silva R, Fatti L, et al. Fitness-related activities and medical claims related to hospital admissions - South Africa, 2006. *Prev Chronic Dis*. 2009;6(4):A120.

Mayer C, Williams B, Wagner EH, et al. Health care costs and participation in a community-based health promotion program for older adults. *Prev Chronic Dis*. 2010;7(2):A38.

Nguyen HQ, Ackermann RT, Berke EM, et al. Impact of a managed-Medicare physical activity benefit on health care utilization and costs in older adults with diabetes. *Diabetes Care*. 2007;30(1):43-8.

Orme-Johnson DW, Herron RE. An innovative approach to reducing medical care utilization and expenditures. *Am J Manag Care*. 1997;3(1):135-44.

Ozminkowski RJ, Goetzel RZ. Getting closer to the truth: overcoming research challenges when estimating the financial impact of worksite health promotion programs. *Am J Health Promot*. 2001;15(5):289-95.

Ozminkowski RJ, Goetzel RZ, Wang F, et al. The savings gained from participation in health promotion programs for Medicare beneficiaries. *J Occup Environ Med*. 2006;48(11):1125-32.

Ozminkowski RJ, Ling D, Goetzel RZ, et al. Long-term impact of Johnson & Johnson's Health & Wellness Program on health care utilization and expenditures. *J Occup Environ Med*. 2002;44(1):21-9.

Patel D, Lambert EV, da Silva R, et al. Participation in fitness-related activities of an incentive-based health promotion program and hospital costs: a retrospective longitudinal study. *Am J Health Promot*. 2011;25(5):341-8.

Patel DN, Lambert EV, da Silva R, et al. The association between medical costs and participation in the vitality health promotion program among 948,974 members of a South African health insurance company. *Am J Health Promot*. 2010;24(3):199-204.

Pelletier KR. A review and analysis of the clinical and cost-effectiveness studies of comprehensive health promotion and disease management programs at the worksite: update VI 2000-2004. *J Occup Environ Med*. 2005;47(10):1051-8.

Pronk NP, Boucher JL, Gehling E, et al. A platform for population-based weight management: description of a health plan-based integrated systems approach. *Am J Manag Care*. 2002;8(10):847-57.

Serxner S, Gold D, Anderson D, et al. The impact of a worksite health promotion program on short-term disability usage. *J Occup Environ Med*. 2001;43(1):25-9.

Shephard RJ, Corey P, Renzland P, et al. The influence of an employee fitness and lifestyle modification program upon medical care costs. *Can J Public Health*. 1982;73(4):259-63.

Shephard RJ, Corey P, Renzland P, et al. The impact of changes in fitness and lifestyle upon health care utilization. *Can J Public Health*. 1983;74(1):51-4.

Shephard RJ, Corey P, Renzland P, et al. Fitness program reduces health care costs. *Dimens Health Serv*. 1982;59(1):14-5.

Spencer LS, Pratt DS, Hausman A. Using health benefits as an incentive to change employee health risks: preliminary program results. *Empl Benefits J*. 1996;21(2):26-9, 36.

Spilman MA, Goetz A, Schultz J, et al. Effects of a corporate health promotion program. *J Occup Med*. 1986;28(4):285-9.

Stevens W, Hillsdon M, Thorogood M, et al. Cost-effectiveness of a primary care based physical activity intervention in 45-74 year old men and women: a randomised controlled trial. *Br J Sports Med*. 1998;32(3):236-41.

Terry PE, Pheley AM. Health risks and educational interests in an HMO. *HMO Pract*. 1991;5(1):3-6.

Wylie-Rosett J, Swencionis C, Ginsberg M, et al. Computerized weight loss intervention optimizes staff time: the clinical and cost results of a controlled clinical trial conducted in a managed care setting. *J Am Diet Assoc*. 2001;101(10):1155-62; quiz 1163-4.

APPENDIX C. DATA ABSTRACTION ELEMENTS

Study Characteristics:

- Study design
- Study dates
- Study setting
- Geographical location
- Funding source
- Subject selection/enrollment in study
- Source of comparator population and comparator strategy
- Patient eligibility criteria for study matching characteristics

Population Characteristics:

- Number of subjects
- Gender
- Race
- Age
- Education
- Baseline physical activity level
- Smoking status
- Weight
- HbA1c (%)
- Lipids (total cholesterol, LDL, HDL)
- Blood pressure (systolic, diastolic)
- Other relevant comorbid conditions and baseline characteristics (e.g., comorbidities, preventive services index, health care use, health care costs, distance to fitness facility, income)

Intervention Components:

- Components of benefit
- Structure of benefit (e.g., vouchers, rebates, premium reductions)
- Payment structure
- Type of fitness centers available through benefit

Outcome Components:

- Physical activity participation rates
- Weight control
- Pain level using validated measures
- Biophysical markers
  - Glucose control
  - Blood pressure control
- Health-related quality of life
- Health care utilization of medical resources
- Health care costs
- Patient satisfaction with health plan
- Retention of plan members

APPENDIX D. PEER REVIEW COMMENTS

Reviewer	Comment	Response
Question 1: Are the objectives, scope, and methods for this review clearly described?		
1	Yes. The scope is clear but the aims are worded as if the focus is on whether benefit packages that include incentives for fitness center use increase physical activity. Clearly the scope is broader than this, as evidenced by the conceptual model and the sub aims of key question #1. There are a multitude of health benefit designs that involve incentives. The goal is more generally to improve health, reduce costs, and to meet a market demand (i.e. a more common reason employer health executives report including incentives is that they are "part of the plan"; in other words, they can't get by designing a health benefit package any longer that leaves health and wellness benefits out of the package – there is a demand among purchasers). The review searched for effects beyond physical activity. I would recommend rewording the key questions to reflect the interest in whether benefits result in improved health when they include incentives to use a fitness facility. This frames the focus on broader metrics of improved health but still defines the focus on fitness facilities, as opposed to the multitude of other health/wellness programs and incentives in the marketplace today. I have to keep reminding myself that the focus is really very specifically on incentives to encourage fitness center use. The "effects" are broader than physical activity (e.g. participation, downstream effects of PA, and cost/utilization). Consider: What are the effects of policy/ benefits packages that include vouchers, rebates, premium reductions, or other economic incentives to improve health specifically through encouraging the use of fitness centers.	Thank you. We agree that the scope is beyond physical activity. The Key Questions, however, were created in collaboration with, and approved by, the technical expert partners and stakeholders. Therefore, we cannot modify them at this time for the report.
2	Yes, and no comments from reviewer 2.	Thank you.
3	Yes, and no comments from reviewer 3.	Thank you.
4	Yes. Clear delineation of purpose and KQs, scope is defined with clear inclusion criteria and exclusion criteria, and the methods follow clearly outlined systematic review processes.	Thank you.
Question 2: Is there any indication of bias in our synthesis of the evidence?		
1	No. I do think that there are likely a large number of "program evaluations" that have been conducted by employers or health plans to ascertain net costs of such a benefit. To the extent that these evaluations are absent from the literature, it likely implies that they were negative studies or of poor quality.	Thank you. We agree, and our focus was on peer-reviewed, published literature.
2	No, and no comments from reviewer 2.	Thank you.
3	No, and no comments from reviewer 3.	Thank you.
4	No. Transparency in logic makes the results of the logic flow consistent with the findings presented	Thank you.
Question 3: Are there any published or unpublished studies that we may have overlooked?		
1	No. I do not know of any. There was a study by Fody-Urias about 10 years ago that looked at SS, but I believe it was uncontrolled and likely of limited methodologic rigor.	Thank you. This study was excluded during citation screening but marked for background. Although this paper did look at Silver Sneakers, the study design was not one of interest because there was no control group.

Reviewer	Comment	Response
2	It would have been useful and relevant to look at or contrast the literature of **employer sponsored programs** to encourage PA through fitness center memberships. Whether an is employer sponsored vs healthplan sponsored seems like a trivial distinction in terms of potential outcomes and mechanism related to effectiveness.	Thank you. Employer-sponsored programs were not in the scope of the project. This project was focused on the VA context (i.e., health plan members as opposed to employees).
3	No, and no comments from reviewer 3	Thank you.
4	No. The search was in-depth and appears to have included the correct type of studies. Personally, I am also not aware of any other research studies that have been reported in this area.	Thank you.
Question 4: Please write additional suggestions or comments below. If applicable, please indicate the page and line numbers from the draft report.		
1	There are a few missed opportunities to frame the discussion for consumption by both the VA and the research community. I have highlighted three of these in comment boxes on the draft. There is also a fourth, related to behavioral economics, which is also discussed below. First, the authors should review and reference the Kaiser Family Foundation annual employer survey. In the most recent year, the survey captured the current practices of >2000 employers regarding the goals and design of health wellness programs and incentives included in health benefit design. Clearly, the train has left the station. To the extent that there may be 'better practices' and resources for some benefit designs may constitute waste, these programs demand greater evaluation.	Thank you. We have added information related to the Kaiser Family Foundation and Health Research and Educational Trust Employer Health Benefits 2012 Annual Survey in the discussion section.
	Second, the issue of distance as a moderator of the effect of fitness center incentives and use should be expanded more and tailored more for the VA. Specifically, give consideration to the challenges of VA facility-based health promotion programs in the context of large VISN areas. Consider the role of "local" programs that are provided closer to a Veteran's home, or at another Veteran service organization that is closer to them. These still need evaluation but could frame new research questions for VA investigators.	Thank you. We expanded on the issue of distance as relates to VA in the discussion section.
	Third, the discussion of alternative study designs is very good but also general. Consider giving an example of a randomized encouragement trial that blends to generalizability of a pragmatic, whole-population evaluation with the internal validity of randomization. See my comment on the draft. Last, I would recommend a slightly deeper discussion of the need for evaluating both the fitness resource as well as the nature of how fitness center use is encouraged. Behavioral econ studies have shown to some extent that you get what you pay for. In other words, if the incentive is for participation, there is greater participation but not necessarily improved health behavior or outcomes. The nature and direction of the incentive is a critical point of study in the future evaluation of these benefits.	Thank you. We elaborated on the role of incentives and encouragement in the discussion section.

Reviewer	Comment	Response
1	There is a missed opportunity here to underscore the enormity of the issue. The horse is out of the barn and we know very little about the health or economic impacts of these forms of incentives, which could divert limited resources from other areas of employee compensation (e.g. salary, other health benefits, lower premiums). Please see the annual Kaiser Family Foundation Employer Survey. The most recent findings are summarized on the kff website and could help to frame the magnitude of efforts to provide such incentives in the commercial sector.	Thank you. See response to related comments above.
	Cash rewards are currently the most common form of incentive used in the large employer sector to encourage wellness program use. It might be listed here separately from "other economic incentives" (under "Intervention") for this purpose. - add blood lipid control and improved work productivity (reduced absenteeism and presenteeism) to possible intermediate outcomes? - add Non-medical cost reductions to Final Outcomes (again economic impact of work productivity, absence, disability, death, and replacement)?	Thank you. We understand these comments are regarding the framework presented. Figure 1 represents an analytic rather than conceptual framework. As such, the analytic framework is not intended to be a comprehensive representation of all outcomes.
1	There is another opportunity here to relate this to the VA. Because many Veterans travel great distances to access VA facilities, a model that provides or encourages access to peripheral fitness facilities or programs (commercial or through CBOCs or other veteran-accessible organizations such as a VFW) could, in fact, have advantages. This may raise the relevance of encouraging greater research in this area as it may help specifically to serve unique needs of veteran patients.	Thank you. We added details on travel distance for Veterans to the discussion section.
	Might discuss briefly CRET designs. Specifically, pragmatic trials that randomize at the employer level to encouragement versus no encouragement of eligible employees to participate in a physical fitness benefit program. Treatment effects are estimated using an instrumental variables approach that uses random assignment as the instrument. This allows simultaneous assessment of both the encouragement step (interventions may include cash rewards, premium discounts, etc) and use of the fitness facility itself.	Thank you. We added information on the use of incentives as a strategy to increase compliance with the fitness center benefit to the Discussion section.
2	Given that only one cohort of patients was found I agree it is very hard to make any generalizable statements regarding KQ1.	Thank you.
	In consideration of this proposed benefit the key piece of information is how likely people who are not likely to be regularly physically active without the benefit will convert to being physically active with the benefit. If people are already going to fitness clubs, addition of the benefit will not be beneficial. Similar the large number of people with no interest in attending fitness clubs lowers the of health improvement to a population who is offered the benefit change. I understand this is out of scope of this review but does highlight the relevant questions for population health improvement.	Thank you. We agree these are important points to consider, which are further elaborated and addressed in the Discussion section.

Reviewer	Comment	Response
3	I believe it would be helpful to look at Health Service Research Centers associated with health plans and review HEDIS measures of plans that offer gym membership for member utilization and satisfaction, and to see if a retrospective study could be possible using a common methodology across plans for behavioral and health outcomes. Niko Pronk of Health Partners would be a good place to start.	Thank you. This is an excellent idea for future study and one of the study designs addressed in the Recommendations for Future Research.
4	On p. 6, in the Conclusion section, the end of the second sentence indicates the evidence remains "weak". I would recommend changes this statement somewhat to read "…remains insufficient and weak, mostly due to study design limitations and small number of studies that meet criteria for inclusion." On p. 8, second paragraph, line 9: I would recommend to insert "full or partial" between the words "providing" and "membership"	Thank you. These changes were incorporated.
Optional Dissemination and Implementation Questions		
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	I believe VA National Center for Health Promotion at Durham VA is exploring linkages to outside facilities as a potential means to expand the reach or impact of the MOVE program. Perhaps this is what instigated this particular review in the first place. The review team should be aware of this, but it is worth mentioning.	Thank you.
2	No comment from reviewer 2.	Thank you.
3	No comment from reviewer 3.	Thank you.
4	Unsure as I am not too familiar with the VA programs	Thank you.
Question 6: Please provide any recommendations on how this report can be revised to more directly address or assist implementation needs.		
1	Please see my comments under #4 above, as I believe these also have relevance to implementation needs. However, it is clear that the most direct interpretation of the review is that we currently know very little about the costs and effectiveness of health benefit designs that attempt to improve health through external facility use. The immediate need for implementation seems to be the design of natural experiments that help address this knowledge gap, both within and outside the VA. The review should clearly be framed for the VA research community (and I think it is already) and also designed to engage those investigators with program designers at local VAs or at the National Center.	Thank you. We agree, and the reviewer mentions some next likely steps.
2	No comment from reviewer 2.	Thank you.
3	To encourage future research among those with shared interests by speaking at meetings such as AHIP, YMCA-USA, and health plan health service researchers with health promotion interests.	Thank you. The ESP Coordinating Center dissemination plan includes some or all of the following as applicable: cyberseminar, email distribution, briefing to VA leaders, and manuscript publication.

Reviewer	Comment	Response
4	From the systematic review perspective, I don't think there is anything missing. To make the report more directly practical, it may be a good idea to add a series of interviews with health plans that have implemented this type of program to find out what additional analyses they may be using to justify the investment internally. Based on such findings, both a justification for doing this or not may be made as well as a more robust set of information to consider the actual program design elements that may make it a good value for the money.	Thank you. Interviews with health plans may be a direction for future research, but primary data collection is beyond the scope of this review.
Question 7: Please provide us with contact details of any additional individuals/stakeholders who should be made aware of this report.		
1	This should be made accessible to the purchasers (employers, Medicare, and state Medicaid administrators) who are either designing health benefits with public monies or are being asked to purchase such benefits from health plans. The purchasers are looking for answers but are also creating a market demand for these programs with little to know information about whether they work, or how to improve them to work best. This creates waste and could be directing limited resources away from other programs that have greater value for the dollar. Commercial health plans may not be concerned because they are profiting already; why support an evaluation that could show something profitable isn't working? Because of the limited scope and limited information to guide policy changes within the review, this review may be difficult to publish in a scientific journal. However, I would encourage the team try to do this and that distribution to CMMI and to the National Business Group on Health be considered strongly.	Thank you. The ESP Coordinating Center dissemination plan includes some or all of the following as applicable: cyberseminar, email distribution, briefing to VA leaders, and manuscript publication.
2	No comment from reviewer 2.	Thank you.
3	No comment from reviewer 3.	Thank you.
4	I think this report would be of interest to AHIP (America's Health Insurance Plans) and ACHP (the Alliance of Community Health Plans)	Thank you. The ESP Coordinating Center dissemination plan includes some or all of the following as applicable: cyberseminar, email distribution, briefing to VA leaders, and manuscript publication.

APPENDIX E. 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.

Allocation concealment

The method by which randomization assignment is concealed from participants and investigators before and during the enrollment process. Common processes are central allocation (telephone or web-based, pharmacy or off-site statistician controlled randomization sequence generation and sequentially numbered, opaque, sealed envelopes. Allocation concealment concentrates on preventing selection and confounding biases, safeguards the assignment sequence *before and until* allocation, and can always be successfully implemented

Case-control study

A retrospective, analytical, observational study often based on secondary data in which the proportion of cases with a potential risk factor are compared to the proportion of controls (individuals without the disease or condition) with the same risk factor. The common association measure for a case-control study is the odds ratio. These studies are commonly used for initial, inexpensive evaluation of risk factors and are particularly useful for rare conditions or for risk factors with long induction periods. Unfortunately, due to the potential for many forms of bias in this study type, case control studies provide relatively weak empirical evidence even when properly executed.

Case report

A description of a single case, typically describing the manifestations, clinical course, and prognosis of that case. Due to the wide range of natural biologic variability in these aspects, a single case report provides little empirical evidence to the clinician. A case report does describe how others diagnosed and treated the condition and what the clinical outcome was.

Case series

A descriptive, observational study of a series of cases, typically describing the manifestations, clinical course, and prognosis of a condition. A case series provides weak empirical evidence because of the lack of comparability unless the findings are dramatically different from expectations. Case series are best used as a source of hypotheses for investigation by stronger study designs, leading some to suggest that the case series should be regarded as clinicians talking to researchers. Unfortunately, the case series is the most common study type in the clinical literature.

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.

Cochran's Q test

A nonparametric statistic to test for differences in intervention effects between studies. Because the test statistic is often underpowered, the threshold for statistically significant differences in intervention effects is often set at $p<0.10$.

Cohort study

A prospective, analytical, observational study based on data, usually primary, from a followup period of a group in which some have had, have, or will have the exposure of interest, to determine the association between that exposure and an outcome. Cohort studies are susceptible to bias by differential loss to followup, the lack of control over risk assignment, and the potential for zero time bias when the cohort is assembled. Because of their prospective nature, cohort studies are stronger than case-control studies when well executed, but they also are more expensive. Because of their observational nature, cohort studies do not provide empirical evidence that is as strong as that provided by properly executed randomized controlled clinical trials.

Companion article

A publication from a trial that is not the article 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 publication.

Confidence interval (CI)

The range in which a particular result (such as a laboratory test) is likely to occur for everyone in the population of interest a specified percentage of the time known as the confidence level or confidence coefficient. It is an interval calculated from a study's observations used to estimate the reliability of the estimate of a parameter. The most common confidence level is 95%. For example, a confidence interval with a 95% confidence level is intended to give the assurance that, if the statistical model is correct, then taken over all the data that *might* have been obtained, the true value of the parameter will be found within the given interval 95% of the time.

Consistency

The extent to which effect size and direction vary within and across studies; inconsistency may be due to heterogeneity across PICOTS.

Cumulative Index to Nursing and Allied Health Literature (CINAHL)

A collection of medical databases of nursing and allied health literature.

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.

Directness

Degree to which outcomes that are important to users of the comparative effectiveness review (patients, clinicians, or policymakers) are encompassed by trial data.

Embase

A database containing bibliographic records with citations, abstracts, and indexing derived from biomedical and pharmacological articles in peer-reviewed journals.

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.

External validity

The extent to which clinical research studies apply to broader populations. A research study has external validity if its results can be generalized to the larger population.

Forest plot

A visual display of information from individual studies in a meta-analysis. A forest plot shows the amount of variation between the results of the studies as well as an estimate of the overall result of all the studies together. A horizontal line represents the 95% confidence interval (CI) of the "effect" observed in the studies.

Full-text review

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

GRADE

Grading of Recommendations Assessment, Development and Evaluation (GRADE), a systematic approach to evaluating the overall body of research evidence and rating the quality of medical evidence and the strength of clinical recommendations.

Health-related quality of life (HRQOL)

Aspects of overall quality of life that can be clearly shown to affect health—either physical or mental health.

I^2

A statistic that describes the percentage (range from 0–100%) of total variation across studies due to heterogeneity between study characteristics rather than due to chance. Heterogeneity is categorized as low, moderate or high based on I^2 values of 25, 50 or 75%, respectively. It is considered an indication of consistency or inconsistency across studies in a meta-analysis.

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.

Intent-to-treat analysis

A method of analyzing results of a randomized controlled trial that includes in the analysis all cases that should have received a treatment regimen but for some reason did not. All cases allocated to each arm of the trial are analyzed together as representing that treatment arm, regardless of whether they received or completed the prescribed regimen.

Interquartile range (IQR)

A measure of the spread of or dispersion within a data set. The IQR is the width of an interval that contains the middle 50 percent of the sample, so it is smaller than the range and its value is less affected by outliers.

Meta-analysis

A way of combining data from many different research studies. A meta-analysis is a statistical process that combines the findings from individual studies.

Meta-regression analyses

An extension of meta-analysis to subgroups that allows the effect of continuous, as well as categorical, characteristics to be investigated if sufficient studies examining the same characteristics may be compared. In principle, it allows the effect of multiple factors to be investigated simultaneously. In meta-regression, the outcome variable is the effect estimate (e.g., a mean difference, etc.). The explanatory variables are characteristics of studies that might influence the size of the intervention effect.

Mixed effects

Statistical models that include both fixed (nonrandom) and random effects.

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-randomized 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.

Odds ratio

A ratio of the odds of having the outcome of interest in a group with a particular exposure, symptom, or characteristic of interest, to the odds of outcome in a group that does not have the exposure/symptom/characteristic. An odds ratio of 1 indicates that the outcome is equally likely to occur in both groups. An odds ratio of 4 indicates that the outcome is 4 times more likely to be present in the group that has the symptom or characteristic of interest, compared with the group that does not have this symptom. When outcomes are infrequent, the odds ratio is a good approximation of the risk ratio.

PICOTS

Population, intervention, comparator, outcome, timing, setting.

Precision

The degree of certainty for estimate of effect with respect to a specific outcome.

Preferred Reporting Items of Systematic Reviews and Meta-Analyses (PRISMA)

An evidence-based minimum set of items for reporting in systematic reviews and meta-analyses.

Probability

The likelihood (or chance) that an event will occur. In a clinical research study, it is the number of times a condition or event occurs in a study group divided by the number of people being studied.

Prospective observational study

A clinical research study in which people who presently have a certain condition or receive a particular treatment are followed over time and compared with another group of people who are not affected by the condition.

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

A type of study that manipulates a variable between two or more groups, but participants are not randomly assigned to groups. Quasi-experimental study designs, such as nonrandomized pre-post studies, 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.

Relative risk (RR)

A comparison of the risk of a particular event for different groups of people. Relative risk is usually used to estimate exposure to something that could affect health. In a clinical research study, the experimental group is exposed to a particular drug or treatment. The control group is not. The number of events in each group is compared to determine relative risk.

Reporting bias

A bias caused by only a subset of all the relevant data being available. The publication of research can depend on the nature and direction of the study results. Studies in which an intervention is not found to be effective are sometimes not published. Because of this, systematic reviews that fail to include unpublished studies may overestimate the true effect of an intervention. In addition, a published report might present a biased set of results (e.g., only outcomes or subgroups where a statistically significant difference was found).

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).

Standard error

The standard deviation of the sampling distribution of a statistic. Measurements taken from a sample of the population will vary from sample to sample. The standard error is a measure of the variation in the sample statistic over all possible samples of the same size. The standard error decreases as the sample size increases.

Standardized mean difference (SMD)

The difference between two estimated means divided by an estimate of the standard deviation. It is used to combine results from studies using different ways of measuring the same concept, e.g. mental health. By expressing the effects as a standardized value, the results can be combined since they have no units.

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.