APPENDIX A. SEARCH STRATEGY

PubMed: 12/14/2015

Set	Terms	Results
1	"Wellness coach"[tiab] OR "wellness coaching"[tiab] OR "health coach"[tiab] OR "health coaching"[tiab] OR "peer coach"[tiab] OR "peer coaching"[tiab]	389
2	coaching[tiab] OR coach[tiab]	4710
3	"Health Education"[Mesh] OR "Health Promotion"[Mesh] OR "Motivational Interviewing"[Mesh] OR "Health Behavior"[Mesh] OR "Health Knowledge, Attitudes, Practice"[Mesh] OR "Counseling"[Mesh] OR "Peer Group"[Mesh] OR "Social Support"[Mesh] OR "Self Care"[Mesh] OR "Patient Education as Topic"[Mesh] OR "Exercise"[Mesh] OR "Exercise Therapy"[Mesh] OR "Weight Loss"[Mesh] OR "Nutrition Therapy"[Mesh] OR "Chronic Disease/prevention and control"[Mesh] OR "Chronic Disease/rehabilitation"[Mesh] OR "health education"[tiab] OR "patient education"[tiab] OR "motivational interviewing"[tiab] OR attitudes[tiab] OR attitude[tiab] OR counseling[tiab] OR "social support"[tiab] OR "psychosocial support"[tiab] OR "self care"[tiab] OR "self-efficacy"[tiab] OR "self management"[tiab] OR "physical activity"[tiab] OR "weight loss"[tiab] OR exercise[tiab] OR fitness[tiab] OR nutrition[tiab]	1153106
4	#2 AND #3	2044
5	#1 OR #4	2119
6	(randomized controlled trial[pt] OR controlled clinical trial[pt] OR randomized[tiab] OR randomised[tiab] OR randomization[tiab] OR randomisation[tiab] OR randomly[tiab] OR trial[tiab] OR groups[tiab] OR "Comparative Study"[Publication Type] OR "Controlled Clinical Trial"[Publication Type] OR Nonrandom[tiab] OR non-random[tiab] OR nonrandomized[tiab] OR non-randomized[tiab] OR nonrandomized[tiab] OR non-randomised[tiab] OR “pre-post”[tiab] OR “post-test”[tiab] OR “post test”[tiab] OR pretest[tiab] OR pre-test[tiab] OR “pre test”[tiab] OR quasi-experiment*[tiab] OR quasiexperiment*[tiab] OR quasirandom*[tiab] OR quasi-random*[tiab] OR quasi-control*[tiab] OR quasicontrol*[tiab] OR (controlled[tiab] AND (trial[tiab] OR study[tiab]))) NOT (animals[mh] NOT humans[mh]) NOT (Editorial[ptyp] OR Letter[ptyp] OR Case Reports[ptyp] OR Comment[ptyp])	2902653
7	#5 AND #6	927
8	#7 limit to 2000 – present	891

Embase: 12/14/2015

Set	Terms	Results
1	'Wellness coach':ti,ab OR 'wellness coaching':ti,ab OR 'health coach':ti,ab OR 'health coaching':ti,ab OR 'peer coach':ti,ab OR 'peer coaching':ti,ab	494
2	Coaching:ti,ab OR coach:ti,ab	6298
3	'health education'/exp OR 'health promotion'/exp OR 'motivational interviewing'/exp OR 'health behavior'/exp OR 'attitude to health'/exp OR 'counseling'/exp OR 'peer group'/exp OR 'social support'/exp OR 'self care'/exp OR 'patient education'/exp OR 'exercise'/exp OR 'kinesiotherapy'/exp OR 'weight reduction'/exp OR 'diet therapy'/exp OR 'chronic disease'/exp/dm_rh,dm_dm,dm_pc OR 'health education':ti,ab OR 'patient education':ti,ab OR 'motivational interviewing':ti,ab OR attitudes:ti,ab OR attitude:ti,ab OR counseling:ti,ab OR 'social support':ti,ab OR 'psychosocial support':ti,ab OR 'self care':ti,ab OR 'self-efficacy':ti,ab OR 'self management':ti,ab OR 'physical activity':ti,ab OR 'weight loss':ti,ab OR exercise:ti,ab OR fitness:ti,ab OR nutrition:ti,ab	1754744
4	#2 AND #3	2946
5	#1 OR #4	3048

Set	Terms	Results
6	'randomized controlled trial'/exp OR 'crossover procedure'/exp OR 'double blind procedure'/exp OR 'single blind procedure'/exp OR random*:ti,ab OR factorial*:ti,ab OR crossover*:ti,ab OR (cross NEAR/1 over*):ti,ab OR placebo*:ti,ab OR (doubl* NEAR/1 blind*):ti,ab OR (singl* NEAR/1 blind*):ti,ab OR assign*:ti,ab OR allocat*:ti,ab OR volunteer*:ti,ab OR 'clinical study'/exp OR 'clinical trial':ti,ab OR 'clinical trials':ti,ab OR 'controlled study'/exp OR (controlled:ti,ab AND (trial:ti,ab OR study:ti,ab)) OR (non NEAR/1 random*):ti,ab OR (quasi NEAR/1 experiment*):ti,ab OR (quasi NEAR/1 random*):ti,ab OR (quasi NEAR/1 control*):ti,ab OR 'comparative effectiveness'/exp OR 'comparative study'/exp OR 'comparative study':ti,ab OR 'comparative studies':ti,ab OR nonrandom*:ti,ab OR 'pre-post':ti,ab OR 'post test':ti,ab OR pretest:ti,ab OR pre-test:ti,ab OR 'pre test':ti,ab OR quasiexperiment*:ti,ab OR quasirandom*:ti,ab OR quasicontrol*:ti,ab NOT ('case report'/exp OR 'case study'/exp OR 'editorial'/exp OR 'letter'/exp OR 'note'/exp) AND [humans]/lim	**6095040**
7	#5 AND #6	**1388**
8	[embase]/lim NOT [medline]/lim	
9	#7 AND #8	**560**
10	#9 limit to 2000 – present	**552**

CINAHL: 12/14/2015

Set	Terms	Results
1	TI ("Wellness coach" OR "wellness coaching" OR "health coach" OR "health coaching" OR "peer coach" OR "peer coaching") OR AB ("Wellness coach" OR "wellness coaching" OR "health coach" OR "health coaching" OR "peer coach" OR "peer coaching")	**305**
2	TI (coaching OR coach) OR AB (coaching OR coach)	**4864**
3	(MH "Health Education+") OR (MH "Health Promotion+") OR (MH "Motivational Interviewing") OR (MH "Health Behavior+") OR (MH "Attitude to Health+") OR (MH "Health Knowledge") OR (MH "Counseling+") OR (MH "Peer Group") OR (MH "Support, Psychosocial+") OR (MH "Social Environment+") OR (MH "Self Care+") OR (MH "Patient Education+") OR (MH "Exercise+") OR (MH "Therapeutic Exercise+") OR (MH "Weight Loss+") OR (MH "Diet Therapy+") OR (MH "Chronic Disease/PC") OR (MH "Chronic Disease/RH") OR TI ("health education" OR "patient education" OR "motivational interviewing" OR attitudes OR attitude OR counseling OR "social support" OR "psychosocial support" OR "self care" OR "self-efficacy" OR "self management" OR "physical activity" OR "weight loss" OR exercise OR fitness OR nutrition) OR AB ("health education" OR "patient education" OR "motivational interviewing" OR attitudes OR attitude OR counseling OR "social support" OR "psychosocial support" OR "self care" OR "self-efficacy" OR "self management" OR "physical activity" OR "weight loss" OR exercise OR fitness OR nutrition)	**612054**
4	#2 AND #3	**2379**
5	#1 OR #4	**2438**

Set	Terms	Results
6	(MH "Clinical Trials+") OR PT Clinical trial OR PT randomized controlled trial OR TX clinic* n1 trial* OR TX ((singl* n1 blind*) OR (singl* n1 mask*)) OR TX ((doubl* n1 blind*) OR (doubl* n1 mask*)) OR TX ((tripl* n1 blind*) OR (tripl* n1 mask*)) OR TX ((trebl* n1 blind*) OR (trebl* n1 mask*)) OR (MH "Random Assignment") OR (MH "Quantitative Studies") OR TI (randomized OR randomised OR randomization OR randomisation OR randomly OR trial OR groups) OR AB (randomized OR randomised OR randomization OR randomisation OR randomly OR trial OR groups) OR TI (Nonrandom OR non-random OR nonrandomized OR non-randomized OR nonrandomized OR non-randomised) OR AB (Nonrandom OR non-random OR nonrandomized OR non-randomized OR nonrandomized OR non-randomised) OR (MH "Comparative Studies") OR (MH "Quasi-Experimental Studies+") OR TI (“pre-post” OR “post-test” OR “post test” OR pretest OR pre-test OR “pre test” OR quasi-experiment* OR quasiexperiment* OR quasirandom* OR quasi-random* OR quasi-control* OR quasicontrol* OR (controlled AND (trial OR study))) OR AB (“pre-post” OR “post-test” OR “post test” OR pretest OR pre-test OR “pre test” OR quasi-experiment* OR quasiexperiment* OR quasirandom* OR quasi-random* OR quasi-control* OR quasicontrol* OR (controlled AND (trial OR study))) NOT (PT editorial OR PT letter OR PT case study OR PT commentary)	**1371759**
7	#5 AND #6	**1288**
8	#7 limit to 2000 – present	**1239**

PsycINFO: 12/14/2015

Set	Terms	Results
1	TI ("wellness coach" OR "wellness coaching" OR "health coach" OR "health coaching" OR "peer coach" OR "peer coaching") OR AB ("wellness coach" OR "wellness coaching" OR "health coach" OR "health coaching" OR "peer coach" OR "peer coaching")	**303**
2	TI (coaching OR coach) OR AB (coaching OR coach)	**11927**
3	DE "Health Education" OR DE "Health Promotion" OR DE "Motivational Interviewing" OR DE "Health Behavior" OR DE "Health Knowledge" OR DE "Health Attitudes" OR DE "Counseling" OR SU counseling OR DE "Peer Relations" OR DE "Peer Pressure" OR DE "Peers" OR DE "Social Support" OR DE "Support Groups" OR DE "Self-Care Skills" OR DE "Client Education" OR DE "Exercise" OR DE "Aerobic Exercise" OR DE "Weight lifting" OR DE "Yoga" OR DE "Movement Therapy" OR DE "Weight Loss" OR DE "Diets" OR TI ("health education" OR "patient education" OR "motivational interviewing" OR attitudes OR attitude OR counseling OR "social support" OR "psychosocial support" OR "self care" OR "self-efficacy" OR "self management" OR "physical activity" OR "weight loss" OR exercise OR fitness OR nutrition) OR AB ("health education" OR "patient education" OR "motivational interviewing" OR attitudes OR attitude OR counseling OR "social support" OR "psychosocial support" OR "self care" OR "self-efficacy" OR "self management" OR "physical activity" OR "weight loss" OR exercise OR fitness OR nutrition)	**472340**
4	#2 AND #3	**2605**
5	#1 OR #4	**2739**

Set	Terms	Results
6	ZC "treatment outcome/clinical trial"OR DE "Clinical Trials" OR TI (randomized OR randomised OR randomization OR randomisation OR randomly OR trial OR trials OR groups) OR AB (randomized OR randomised OR randomization OR randomisation OR randomly OR trial OR trials OR groups) OR TI ("comparative study") OR AB ("comparative study") OR TI (Nonrandom OR non-random OR nonrandomized OR non-randomized OR nonrandomized OR non-randomised OR “pre-post” OR “post-test” OR “post test” OR pretest OR pre-test OR “pre test” OR quasi-experiment* OR quasiexperiment* OR quasirandom* OR quasi-random* OR quasi-control* OR quasicontrol* OR (controlled AND (trial OR study)))) OR AB (Nonrandom OR non-random OR nonrandomized OR non-randomized OR nonrandomized OR non-randomised OR “pre-post” OR “post-test” OR “post test” OR pretest OR pre-test OR “pre test” OR quasi-experiment* OR quasiexperiment* OR quasirandom* OR quasi-random* OR quasi-control* OR quasicontrol* OR (controlled AND (trial OR study)))) AND (ZZ "journal article")	**877890**
7	#5 AND #6	**893**
8	#7 limit to 2000 – present; limit to human	**784**

APPENDIX B. STUDY CHARACTERISTICS TABLE

For full study citations, please refer to the report's main reference list.

Study Country Veteran?	Condition Setting Duration	Intervention Strategies Used	Theoretical Orientation Therapeutic Model	Comparator	Outcomes Abstracted	Risk of Bias Funding Source
Annesi, 2011[22] USA No	Obesity Community 6 months	Six 1-hour individual meetings with YMCA "Coach Approach" trained wellness specialist + at-home exercise prescription (3/week for 24 weeks for total of 72 sessions) Goal-setting, self-monitoring, and chronic disease education	Social cognitive theory Self-efficacy (Bandura)	6 individual meetings with standard- trained fitness specialist + 72 at-home exercise sessions Problem-solving, structured, supervised exercise, chronic disease education	Self-efficacy	High YMCA
Appel, 2011[23] USA No	Obesity Primary care 96 weeks (everyone had in-person baseline and end-of-treatment measures)	Coaching in-person (group/individual) weekly for first 12 weeks, monthly (group/individual) next 12 weeks, then either in-person or phone for last 72 weeks by trained, supervised health professional + website and email Goal-setting, self-monitoring, problem-solving, chronic disease education, and "learning modules online"	Social cognitive theory Behavioral self-management Motivational interviewing	(1) Coaching support delivered remotely by phone, study-specific website, and email (2) Self-directed weight loss using website (baseline and 96-week follow-up)	Weight change BMI	Unclear NIH: NHLBI
Blackberry, 2013[24] Australia No	Type 2 diabetes Primary care 18 months	1 in-person baseline assessment, then 8 structured phone sessions on self-management of diabetes with coaching by trained, supervised general practice nurse; written session summaries provided to patient and primary care physician Self-monitoring and "coaching on patient-provider communication"	NR NR	After 1 in-person baseline assessment, usual care was provided including referrals to dieticians and other diabetes specialists	A1c Weight change Self-efficacy	Low Australian National Health and Medical Research Council

Study Country Veteran?	Condition Setting Duration	Intervention Strategies Used	Theoretical Orientation Therapeutic Model	Comparator	Outcomes Abstracted	Risk of Bias Funding Source
Bostrom, 2016[25] Sweden No	Systemic lupus erythematosus Rheumatology clinic 12 months	(1) 0-3 months: Individual, in-person 1-hour coaching by physiotherapist at study start, 6 weeks, and 12 weeks; general education, supervised aerobic exercise, loan and use of heart rate monitor, and use of physical activity diary (2) 4-12 months: Some physical activity supervision, heart rate monitor, and diary None	Social cognitive theory Behavior theory model	Usual care at rheumatology clinic, but patients in control group were asked not to change their activity level during the study	Physical activity	Low Swedish Rheumatism Association and Vardal Foundation, Karolinska (Univ) Institute
Brodin, 2008[26] Sweden No	Rheumatoid arthritis Rheumatology clinics 12 months	Phone support after 1 week, moving to once monthly by physical therapist coach; physical function tests every 3 months to encourage adherence to graded activity goals, feedback given Goal-setting, problem-solving, chronic disease education	Cognitive behavioral therapy NR	Usual care (no description given other than “control group”)	Physical activity	High Government, Swedish Research Council, the Vardal Foundation, the Swedish Rheumatism Association
Browning, 2014[27] China No	Type 2 diabetes Community Health Center 12 months	Heath coaching by nurse via in-person + phone (both 2/month for first 3 months) diminishing over next 9 months; maximum contact was 19 phone and 18 in-person sessions Not reported	Transtheoretical model/ stages of change Motivational interviewing	Usual care provided by family physician where patients were typically referred to diabetes specialists or to Traditional Chinese Medicine practitioners	A1c BMI	Unclear Government and private foundation
Cinar, 2014[28] Turkey No	Type 2 diabetes Hospital clinics 13 months	In addition to standard health education, 2 in-person individual visits + single 10- to 20-minute phone call within first 3 weeks; 1 in-person + 1 call in next 6 months; 1 in-person + 1 call in last 6 months, for up to 7 total contacts with the behavioral health specialist coach Self-monitoring, chronic disease education	NR NR	Health education consisting of 3 seminars on oral health and diabetes management	A1c	Unclear Government, International Research Fund

Study Country Veteran?	Condition Setting Duration	Intervention Strategies Used	Theoretical Orientation Therapeutic Model	Comparator	Outcomes Abstracted	Risk of Bias Funding Source
Damschroder, 2014[29] USA Yes	Obesity VA medical centers 12 months	ASPIRE-Group: Coaching via in-person 90-minute group sessions with a specially trained lifestyle coach for 1/week for 3 months, then 2/month for 6 months, then 60-minute sessions 1/month for last 3 months Goal-setting, self-monitoring, problem-solving, chronic disease education	Unclear Problem-solving therapy	(1) ASPIRE-Phone: Coaching via phone for 30 minutes, 1 time/week for 3 months, then 20 minutes for remaining time (2 times/month for 6 months decreasing to 1 time/month for last 3 months) (2) Standard VA MOVE! program	Weight change BMI Physical activity Diet	High VA
Frosch, 2011[21] USA No	Type 2 diabetes Primary care Duration NR	Phone coaching by trained nurse diabetes educator, 5 sessions total: first session for 60 minutes; sessions 2-3 for 30 minutes, sessions 4-5 for 15 minutes Goal-setting, self-monitoring, problem-solving, chronic disease education	NR Motivational interviewing	Education brochure on diabetes management; no other strategies employed	A1c BMI Physical activity Diet Medication adherence	Unclear NIA/NIH, private foundation
Glasgow, 2003[30] USA No	Type 2 diabetes Primary care 10 month (40 weeks)	Internet-based basic information + either (1) tailored self-management (computer-mediated access to trained professional coach approximately twice weekly or (2) peer support via online forum and newsletters Goal-setting, self-monitoring, chronic disease education	Self-efficacy theory NR	In-home training to use website providing chronic disease education without additional support	A1c Physical activity Diet	High NIH: NIDDK
Hawkes, 2013[31] Australia No	Colorectal cancer Cancer registry 6 months	11 health coaching sessions biweekly for 5 months via phone by nurse, behavioral specialist, or health educator (average duration of call, 31.5 minutes) + handbook + motivational postcards + pedometer Goal-setting, self-monitoring, chronic disease education	NR Acceptance and commitment therapy	Usual care + educational brochures on understanding colorectal cancer, cutting cancer risk, diet, and physical activity + quarterly mailed educational newsletter	BMI Physical activity Diet Smoking	Unclear Australian government (cancer division of health branch)

Study Country Veteran?	Condition Setting Duration	Intervention Strategies Used	Theoretical Orientation Therapeutic Model	Comparator	Outcomes Abstracted	Risk of Bias Funding Source
Holland, 2005[32] USA No	Mixed: at least one chronic condition; unspecified Community 12 months	In-person meeting with nurse at baseline and 6 months, minimum 4 health coaching calls in between, 12 monthly newsletters, fitness program Goal-setting, chronic disease education, counseling with MSW, if depressed	NR NR	Usual care; controls were not recontacted by the program until the anniversary date of their initial interview for follow-up	BMI Physical activity	Unclear Private foundation
Karhula, 2015[33] Finland No	Mixed population (type 2 diabetes and CVD) Community integrative care 12 months	One coaching phone call from employee trained in Pfizer coaching model every 4-6 weeks (target=12 total); length of call approximately 30 minutes and emphasized problem-solving skills + monitoring of weight, blood glucose, SBP, and/or step count dependent on diagnosis via mobile application Problem-solving, self-monitoring	Wagner's chronic care Pfizer's health coaching model	Usual care; no further details or description of control group given	A1c (diabetes only) Weight (diabetes and CVD separately)	Unclear Government, European Commission, Industry
Kim, 2015[34] USA No	Type 2 Diabetes Community 13-14 months	Six 2-hour group sessions over 6 weeks, then monthly coaching calls for 1 year from trained nurses or community health workers; calls ranged 15-45 minutes Goal-setting, self-monitoring, problem-solving, chronic disease education	Precede-Proceed Motivational interviewing, problem-solving therapy	Waitlist; no further details given other than control was oversampled to assure adequate retention	A1c Self-efficacy	High NIH: NIDDK
Knittle, 2015[35] Netherlands No	Rheumatoid arthritis Specialty clinics 18 weeks	2 in-person, individual coaching sessions with rheumatology nurse, 40-60 minutes, at weeks 4 and 5; 3 followup phone calls, 20 minutes, weeks 6, 12, and 18 Goal-setting, self-monitoring, problem-solving, chronic disease education	Health Belief Model, self-regulation theory Motivational interviewing, problem-solving therapy, self-regulation theory	Education via 1 in-person group session with nurse	Physical activity Self-efficacy Functional status	Unclear Private foundation

Study Country Veteran?	Condition Setting Duration	Intervention Strategies Used	Theoretical Orientation Therapeutic Model	Comparator	Outcomes Abstracted	Risk of Bias Funding Source
Lin, 2013[16] USA No	Hypertension Primary care 18 months total (5 months group intervention)	Weekly small groups for 20 weeks with trained health educators and dieticians + manual; strategies to manage weight and blood pressure via DASH diet, increase exercise; and coaching strategies; during and after group intervention, a peer educator phoned participants monthly for a total of 18 calls Goal-setting, self-monitoring, problem-solving, chronic disease education	Transtheoretical Model / Stages of Change; Social-Cognitive Theory Motivational interviewing, problem-solving therapy	Usual care was an individual visit with interventionist to receive advice + written materials on lifestyle modification for blood pressure control consistent with JNC-7 guidelines	SBP Diet	Low NIH
Luley, 2014[36] Germany No	Metabolic syndrome Community setting 12 months	1 basic education session + monthly health coaching call from trained physician or nurse, each approximately 20 minutes + accelerometer (data transmitted to coach as basis for phone calls) None reported	NR NR	After 1 basic education session that included an explanation of importance of physical activity and diet, patients were randomized, then control group left	BMI	High German Federal Ministry of Education and Research
Ma, 2013[37] (Companion study, Azar, 2013[38]) USA No	Obesity Primary care 15 months (60 weeks)	Lifestyle Balance of 2 weekly, in-person group sessions (90-120 minutes) using goal-setting, with food tastings and 30-45 minutes of guided exercise led by coach-dietician followed by 12-month maintenance phase; personalized email from coach monthly Goal-setting, self-monitoring, problem-solving, chronic disease education, structured exercise, relapse prevention	NR NR	(1) Self-led via DVD and email correspondence with coach/RD that used goal-setting, self-monitoring, and chronic disease education (2) Usual care; no further details given	Weight BMI	Low NIH: NIDDK, private foundation

Study Country Veteran?	Condition Setting Duration	Intervention Strategies Used	Theoretical Orientation Therapeutic Model	Comparator	Outcomes Abstracted	Risk of Bias Funding Source
McMurray, 2002[39] USA No	End stage renal disease + diabetes Dialysis unit 1 year (52 weeks)	Minimum of monthly (for peritoneal patients) in-person, individual sessions with diabetes care manager for motivational coaching; weekly contact as needed by phone with manager, social worker, registered dietician, or registered nurse to cover self-management and diabetes care; maximum of 3 times/week (for hemodialysis patients) Problem-solving	NR NR	Usual care at a standard dialysis unit	A1c	Unclear National Kidney Foundation
Nishita, 2013[40] USA No	Type 2 diabetes Community workplace setting 12 months	Average of ten 1-hour in-person, individual sessions with certified health coach and four 45-minute sessions with pharmacist over intervention year Goal-setting, self-monitoring, problem-solving, chronic disease education	Health belief model, self-determination theory Motivational interviewing, problem-solving therapy	Usual care; no further details given	A1c BMI Self-efficacy	Unclear Centers for Medicare and Medicaid Services
Patja, 2012[41] Finland No	Mixed: Type 2 diabetes, CVD, CHF Primary care and hospital 12 months	Monthly phone calls with nurse coach (initial duration averaging 60 minutes, decreasing to 30 minutes over time); call completion averaged 10-11 calls over year; optional followup calls were rarely utilized Goal-setting, self-monitoring, chronic disease education	Self-regulation theory Motivational interviewing	Usual care; article states “control arm” and, with no other details given, usual care is assumed because of recruitment sites used	A1c (diabetes only) SBP (CVD only)	High Government: Finland Innovation Fund, industry
Pearson, 2013[19] (Companion study, Pearson 2012[42]) Canada No	Obesity University 12 weeks	Phone coaching sessions with certified health coach 1 time/week for 12 weeks; average length of call was 45 minutes Goal-setting, problem-solving	NR Motivational Interviewing and CBT	Scripted education-based phone lessons using cognitive behavioral therapy principles from LEARN manual 1/week for 12 weeks; average length of call was 30-45 minutes Goal-setting, self-monitoring, problem-solving, social support and chronic disease management	Weight change Diet	High Social Sciences and Humanities Research Council of Canada

Study Country Veteran?	Condition Setting Duration	Intervention Strategies Used	Theoretical Orientation Therapeutic Model	Comparator	Outcomes Abstracted	Risk of Bias Funding Source
Pinto, 2015[43] USA No	Breast cancer Community, private practices and hospitals 12 weeks	Health coaching by peer educator via phone 1/week for 12 weeks; average call length was 18 minutes + pedometer + heart rate monitor + physical activity tipsheets Goal-setting, self-monitoring, problem-solving, chronic disease education	Transtheoretical model/ stages of change, social cognitive theory NR	Attention control: phone contact with peer educator 1/week for 12 weeks, but topics centered on breast cancer, not physical activity	Physical activity	High NIH
Ruggiero, 2010[44] USA No	Type 2 diabetes Primary care 6 months	2 in-person, individual contacts (<30 minutes) with certified medical assistant trained in diabetes self-care coaching at baseline and 3 months + 4 monthly phone contacts (<15 minutes) in between clinic visits Goal-setting, self-monitoring, problem-solving, chronic disease education	Transtheoretical model NR	Usual care with physician + basic diabetes education handbook developed by health system staff	A1c	Unclear NIA, NIH
Ruggiero, 2014[20] USA No	Type 2 diabetes Primary care 12 months	Quarterly in-person, individual coaching sessions with specially trained certified medical assistants for 30 minutes at clinic appointments; up to 8 monthly phone calls, 15 minutes, between in-person contacts Goal-setting, self-monitoring, chronic disease education	Transtheoretical model/ stages of change, empowerment theory Motivational interviewing	Enhanced treatment as usual; quarterly physician check-ups; referrals to specialty care (eg, podiatrist, endocrinologist) when necessary; basic education provided by "Diabetes: You're in Control" educational booklet	A1c Diet Physical Activity	High NIH
Sacco, 2009[45] USA No	Type 2 diabetes Primary care 6 months	Coaching call weekly for 3 months (from supervised psychology undergraduate), then every other week for 3 months; average duration of initial call was 54 minutes decreasing to 15-20 minutes Goal-setting, self-monitoring, problem-solving	Social cognitive theory Problem-solving therapy	Control group received treatment as usual from a board-certified endocrinologist	A1c Physical activity Diet Self-efficacy	Unclear Private foundation

Study Country Veteran?	Condition Setting Duration	Intervention Strategies Used	Theoretical Orientation Therapeutic Model	Comparator	Outcomes Abstracted	Risk of Bias Funding Source
Safford, 2015[46] USA No	Type 2 diabetes Primary care 40 weeks	1-hour group diabetes education class + one 5-10 minute individual counseling session to go over baseline "diabetes report cards," then peer coaches phoned weekly for the first 2 months and at least monthly for the next 8 months Goal-setting, self-monitoring, chronic disease education	Health belief model, social cognitive theory and chronic care model NR	1-hour group diabetes education class + 5-10 minute counseling session on a "diabetes report card" showing baseline labs at enrollment	A1c BMI	High American Academy of Family Physicians
Sandroff, 2014[47] USA No	Multiple sclerosis National registry and databases from previous studies over past 5 years 6 months	Coaching (discipline of coach not reported) via internet and 15, one-on-one video sessions (*eg*, Skype) for 6 months decreasing in frequency over time (from weekly to monthly) Goal-setting, self-monitoring, problem-solving	Social cognitive theory NR	Waitlist	Physical Activity Functional status	High National Multiple Sclerosis Society
Sherwood, 2010[48] USA No	Obesity Community and university 20 weeks	DIAL: 2 active arms (same intervention for different durations: 10 sessions or 20 sessions) providing weekly telephone calls with coach (discipline not reported) lasting about 10-20 minutes + pedometer + logbook; calls followed a prescribed sequence in study manual adapted to fit into 10 or 20 lessons None reported	NR NR	Self-directed program participants were sent copy of manual, pedometer, and logbook but were not recontacted until time for follow-up measures	Weight change (kg) Physical activity	High Government grant
Thom, 2013[49] (Companion study, Moskowitz, 2103[50]) USA No	Type 2 diabetes Primary care 6 months (26 weeks)	12-14 sessions of coaching by a peer educator (individual or phone at discretion of subject) with goals of phone contact at least twice/month and 2 or more in-person contacts over 6 months Goal-setting, self-monitoring	NR NR	Usual care included all services normally available, including a nutritionist and diabetes educator via referral from their primary care physician	A1c BMI	unclear Private foundation

Study Country Veteran?	Condition Setting Duration	Intervention Strategies Used	Theoretical Orientation Therapeutic Model	Comparator	Outcomes Abstracted	Risk of Bias Funding Source
Turner, 2012[51] USA No	Hypertension Primary care 6 months (26 weeks)	Phone calls every other month at 1, 3, and 5 months (duration not reported); on alternate months (2 and 4), office-based, in-person, individual counseling sessions (15-30 minutes each) with a peer educator as coach Goal-setting, self-monitoring, problem-solving	Theory of planned behavior Motivational interviewing, problem-solving therapy	Usual care at urban academic general medicine practices	SBP 4-year Framingham Score	Low Private foundation
Vale, 2002[52] Australia No	CAD/CVD NR (most likely cardiology) 6 months (24 weeks)	5 coaching phone calls from dietician, with first call within 2 weeks of randomization; then 3 calls, one every 6 weeks; the fifth call at 24 weeks (to schedule the 6-month assessment); duration of calls varied Self-monitoring, chronic disease education	NR NR	Usual care; no further details given	Total cholesterol	Unclear Industry
Vale, 2003[53] Australia No	CAD/CVD Specialist clinic: cardiology 6 months	5 coaching phone calls from nurse or dietician, with first call within 2 weeks of randomization; then 3 calls, one every 6 weeks; the fifth call at 24 weeks (to schedule the 6-month assessment); duration of calls varied Self-monitoring, chronic disease education	NR NR	Usual care; no further details given	SBP Weight change BMI Diet Smoking	Low Private foundation, industry
Van der Wulp, 2012[54] Netherlands No	Type 2 diabetes Primary care 3 months	3 in-person, individual health coaching sessions, monthly, with trained peer educator using goal-setting; duration of session not reported Goal-setting	Social cognitive theory Motivational interviewing	Usual care from general practitioner based on the Dutch guidelines for type 2 diabetes	Self-efficacy Physical activity	Unclear Private foundation
Varney, 2014[55] Australia No	Type 2 diabetes Diabetes clinic 6 months	Initial coaching call within 2 weeks of randomization followed by at least monthly phone calls (range 4-9 sessions) from dietician coach; duration average 45 minutes initially, then 20 minutes for follow-up calls Goal-setting, self-monitoring, problem-solving	NR Problem-solving therapy	Control group accessed usual care services, including a diabetes clinic staffed by endocrinologists, diabetes educators, and dietitians; patients typically attend the clinic at least monthly, with general practitioner visits occurring as needed	A1c Weight (kg) BMI Physical activity	High Private foundation

Study Country Veteran?	Condition Setting Duration	Intervention Strategies Used	Theoretical Orientation Therapeutic Model	Comparator	Outcomes Abstracted	Risk of Bias Funding Source
Wadden, 2011[56] USA No	Obesity Primary care 24 months (104 weeks)	(1) Coaching only: primary care visits plus 10-15 minute in-person, individual coaching sessions; 2 during the first month, then monthly for 11 months with a trained medical assistant; in months 13-24, coaching could be done by phone every other month (2) Enhanced coaching: as above + choice of meal replacements or weight loss medication Goal-setting, self-monitoring	NR NR	Usual care consisting of quarterly PCP visits that included education about weight management for 5-7 minutes each visit	BMI Weight loss	Unclear NIH: NHLBI
Wayne, 2015[57] Canada No	Type 2 Diabetes Primary care 6 months (26 weeks)	Weekly health coach sessions + exercise education program with smartphone wellness mobile application; components included support for health goals and goal achievement; self-monitoring; discussion of meals, exercise, blood glucose and mood; duration of session 37 (±22) minutes/week; also health coach co-monitored patient's input to mobile application Goal-setting, self-monitoring, structured exercise	NR NR	Weekly health coach sessions + exercise education program without smartphone application; components included support for health goals and goal achievement; self-monitoring; discussion of meals, exercise, blood glucose, and mood; session duration 39 (±28) minutes/week	A1c BMI Weight (kg)	Unclear Government
Whittemore, 2004[58] USA No	Type 2 Diabetes Outpatient diabetes education center 6 months	6 in-person, individual coaching sessions with a trained nurse: first 3 every 2 weeks; then 2 monthly; last session 3 months after first 5 sessions with phone contacts in between sessions 5 and 6 Goal-setting, self-monitoring, problem-solving, chronic disease education	NR Problem-solving therapy	Standard diabetes care, defined as regular visits with a primary care physician at approximately 3- to 4-month intervals; all women randomized to the control condition were invited to participate in the nurse-coaching intervention at the end of the study	A1c BMI Diet Physical Activity	High NIH: NINR

Study Country Veteran?	Condition Setting Duration	Intervention Strategies Used	Theoretical Orientation Therapeutic Model	Comparator	Outcomes Abstracted	Risk of Bias Funding Source
Willard-Grace, 2015[59] (Companion study, Thom, 2015[60]) USA No	Mixed: diabetes, hypertension, elevated lipids Primary care 12 months	5 in-person, individual coaching sessions at baseline, 3, 6, 9, and 12 months with a trained medical assistant as well as monthly follow-ups by phone; total 16 sessions Goal-setting, self-monitoring	NR NR	Patients randomized to usual care had access to any resources available at the clinic, including visits with their clinician, diabetes educators, nutritionists, chronic care nurses, and educational classes	A1c SBP LDL Medication adherence	Unclear Private foundation
Wolever, 2010[61] USA No	Type 2 diabetes Community & registry 22 weeks (5-6 months)	8 calls weekly for first 2 months, then 4 calls biweekly for 2 months; final call 1 month later for total of 14, 30-minute sessions with a trained social worker or medical assistant in psychology coach Goal-setting, problem-solving, chronic disease education	NR Motivational interviewing, mindfulness	Usual care; randomized to the control group received no materials or correspondence during the 6-month period	A1c Medication adherence Physical activity	Unclear Industry
Young, 2014[62] USA No	Type 2 Diabetes Primary care and community Timing unclear: 9-18 weeks	1 in-person, individual session with a nurse coach followed by 5 health coaching sessions via phone or video-conferencing, about once every 2 weeks; average duration of sessions was 30 minutes Goal-setting, self-monitoring	NR Motivational interviewing	Usual care consisted of the services and care available at the rural clinic where the participant received healthcare	Self-efficacy	Unclear NIH: NIDDK, NCATS

Abbreviations: ANCOVA=analysis of covariance, ANOVA=analysis of variance, BMI=body mass index, CI=confidence interval, A1c=glycosylated hemoglobin, JNC=Joint National Committee on Prevention, LDL=low-density lipoprotein-cholesterol., MD=mean difference, MI-via-CALC=Motivational Interview via Co-Active Life Coaching, NCATS=National Center for Advancing Translational Sciences, NHLBI=National Heart, Lung, and Blood Institute, NIDDK=National Institute for Diabetes and Digestive and Kidney Diseases, NIH=National Institutes of Health, NINR=National Institute of Nursing Research, NR=not reported, SBP=systolic blood pressure, SE=standard error, SMD=standardized mean difference

APPENDIX C. STUDY QUALITY ASSESSMENT

RANDOMIZED CONTROLLED TRIALS

Detailed guidance on assessing the risk of bias is found in Higgins J, Altman DG. Chapter 8: Assessing risk of bias in included studies. In Cochrane Handbook for Systematic Reviews of Interventions Version 5.0, 2008. Available at: http://handbook.cochrane.org/chapter_8/8_assessing_risk_of_bias_in_included_studies.htm.

General instructions: Rate each risk of bias item listed below as "Low," "High," or "Unclear."

Rating of individual items:

1.Selection bias

Domain: Random sequence generation

*(**Support for judgement:** Describe the method used to generate the allocation sequence in sufficient detail to allow an assessment of whether it should produce comparable groups.)*

Was the allocation sequence adequately generated?

Low risk High risk Unclear risk

Domain: Allocation concealment?

***(Support for judgement**: Describe the method used to conceal the allocation sequence in sufficient detail to determine whether intervention allocations could have been foreseen in advance of, or during, enrolment)*

Was allocation adequately concealed?

Low risk High risk Unclear risk

Comment

2a. Performance bias (of ONE primary clinical outcome)

Domain: Blinding of participants and "treating" personnel - i.e. the person(s) delivering the intervention.

***(Support for judgement:** Describe all measures used, if any, to blind study participants and personnel from knowledge of which intervention a participant received. Provide any information relating to whether the intended blinding was effective.)*

Was knowledge of the allocated intervention adequately prevented during the study?

Low risk High risk Unclear risk Outcome NR

2b. Performance bias (Medication adherence)

Domain: Blinding of participants and personnel

***(Support for judgement:** Describe all measures used, if any, to blind study participants and personnel from knowledge of which intervention a participant received. Provide any information relating to whether the intended blinding was effective.)*

Was knowledge of the allocated intervention adequately prevented during the study?

Low risk High risk Unclear risk Outcome NR

2b. Performance bias (Physical activity)

Domain: Blinding of participants and personnel

***(Support for judgement:** Describe all measures used, if any, to blind study participants and personnel from knowledge of which intervention a participant received. Provide any information relating to whether the intended blinding was effective.)*

Was knowledge of the allocated intervention adequately prevented during the study?
Low risk High risk Unclear risk Outcome NR

2b. Performance bias (Diet)

Domain: Blinding of participants and personnel
***(Support for judgement:** Describe all measures used, if any, to blind study participants and personnel from knowledge of which intervention a participant received. Provide any information relating to whether the intended blinding was effective.)*

Was knowledge of the allocated intervention adequately prevented during the study?
Low risk High risk Unclear risk Outcome NR

2b. Performance bias (Smoking)

Domain: Blinding of participants and personnel
***(Support for judgement:** Describe all measures used, if any, to blind study participants and personnel from knowledge of which intervention a participant received. Provide any information relating to whether the intended blinding was effective.)*

Was knowledge of the allocated intervention adequately prevented during the study?
Low risk High risk Unclear risk Outcome NR

2c. Performance bias (Self efficacy)

Domain: Blinding of participants and personnel
***(Support for judgement:** Describe all measures used, if any, to blind study participants and personnel from knowledge of which intervention a participant received. Provide any information relating to whether the intended blinding was effective.)*

Was knowledge of the allocated intervention adequately prevented during the study?
Low risk High risk Unclear risk Outcome NR

Comment

3a. Detection bias (of ONE primary clinical outcome):

Domain: Blinding of outcome assessment
***(Support for judgement:** Describe all measures used, if any, to blind outcome assessors from knowledge of which intervention a participant received. Provide any information relating to whether the intended blinding was effective.)*

Was knowledge of the allocated intervention adequately prevented from outcome assessors?
Low risk High risk Unclear risk Outcome NR

3b. Detection bias (Health behavior.):

Domain: Blinding of outcome assessment
***(Support for judgement:** Describe all measures used, if any, to blind outcome assessors from knowledge of which intervention a participant received. Provide any information relating to whether the intended blinding was effective.)*

Was knowledge of the allocated intervention adequately prevented from outcome assessors?
Low risk High risk Unclear risk Outcome NR

Comment

3c. Detection bias (Self efficacy):
Domain: Blinding of outcome assessment
(Support for judgement: *Describe all measures used, if any, to blind outcome assessors from knowledge of which intervention a participant received. Provide any information relating to whether the intended blinding was effective.)*
Was knowledge of the allocated intervention adequately prevented from outcome assessors?
Low risk High risk Unclear risk Outcome NR

Comment

4. Attrition bias:
Domain: Incomplete outcome data
(Support for judgement: *Describe the completeness of outcome data for each main outcome, including attrition and exclusions from the analysis. State whether attrition and exclusions were reported, the numbers in each intervention group (compared with total randomized participants), reasons for attrition/exclusions where reported, and any re-inclusions in analyses performed by the review authors.)*

Were incomplete outcome data adequately addressed?
Low risk High risk Unclear risk

Comment

5. Reporting bias:
Domain: Selective outcomes reporting
*(**Support for judgement:** State how the possibility of selective outcome reporting was examined by the review authors, and what was found.)*

Are reports of the study free of suggestion of selective outcome reporting? (i.e., the author states they will measure an outcome but do not report it)
Low risk High risk Unclear risk

Comment

6. Other
Domain: Other sources of bias
(Support for judgement: *State any important concerns about bias not addressed in the other domains in the tool.*
If particular questions/entries were pre-specified in the review's protocol, responses should be provided for each question/entry.)

Are reports of the study free from other bias due to problems not covered above?
Low risk High risk Unclear risk

Comment

Overall risk of bias rating
Low Unclear High

Narrative:

Risk of Bias	Interpretation	Criteria
Low risk of bias	Bias, if present, is unlikely to alter the results seriously	Adequacy of random sequence generation, allocation concealment, and blinding scored as “low risk of bias” and no important concerns related to the other domains.
Unclear risk of bias	A risk of bias that raises some doubts about the results	One or two domains are scored “not clear” or not done.
High risk of bias	Bias may alter the results seriously	More than 2 domains are scored as “not clear” or not done

* Items contained in Cochrane Risk of Bias Tool

QUALITY ASSESSMENT RESPONSE TABLE

For full study citations, please refer to the report's main reference list.

Study[a]	1a	1b	2a	2b	2b	2b	2b	2c	3a	3b	3c	4	5	6	Overall Risk of Bias Rating
Annesi 2011[22]	UR	UR	NR	NR	NR	NR	NR	UR	NR	NR	UR	LR	LR	LR	HR
Appel 2011[23]	LR	UR	LR	NR	NR	NR	NR	NR	LR	NR	NR	LR	LR	LR	UR
Blackberry 2013[24]	LR	LR	LR	NR	NR	NR	NR	LR	LR	LR	LR	LR	LR	LR	LR
Bostrom 2016[25]	LR	LR	NR	NR	HR	NR	NR	NR	NR	LR	NR	UR	LR	LR[b]	LR
Brodin 2008[26]	LR	HR	NR	NR	HR	NR	NR	NR	NR	LR	NR	HR	LR	LR	HR
Browning 2014[27]	UR	UR	LR	NR	NR	NR	NR	NR	LR	NR	NR	LR	LR	LR[c]	UR
Cinar 2014[28]	UR	LR	LR	LR	UR	UR	UR	UR	LR	LR	LR	HR	LR	LR	UR
Damschroder 2014[29]	LR	UR	HR	NR	HR	HR	NR	NR	HR	HR	NR	LR	LR	LR	HR
Frosch 2011[21]	LR	LR	LR	HR	HR	HR	NR	NR	LR	HR	NR	UR	LR	LR	UR
Glasgow 2003[30]	UR	UR	LR	NR	UR	UR	NR	NR	LR	UR	NR	LR	HR	HR	HR
Hawkes 2013[31]	LR	LR	UR	NR	UR	UR	UR	NR	UR	UR	NR	LR	LR	LR	UR
Holland 2005[32]	LR	UR	UR	NR	HR	NR	NR	NR	UR	UR	NR	UR	LR	LR	UR
Karhula 2015[33]	LR	LR	LR	NR	NR	NR	NR	LR	LR	LR	LR	UR	UR	UR	UR
Kim 2015[34]	HR	HR	LR	NR	NR	NR	NR	HR	LR	NR	HR	HR	LR	UR	HR
Knittle 2015[35]	LR	LR	NR	NR	HR	NR	NR	HR	NR	LR	LR	LR	LR	LR	UR
Lin 2013[16]	LR	LR	LR	NR	NR	LR	NR	NR	LR	LR	NR	LR	LR	LR	LR
Luley 2014[36]	UR	UR	LR	NR	HR	HR	NR	NR	LR	LR	NR	LR	LR	LR	HR
Ma 2013[37] (Azar 2013[38])	LR	LR	LR	NR	NR	NR	NR	NR	UR	NR	NR	LR	LR	LR	LR
McMurray 2002[39]	LR	UR	UR	UR	NR	NR	NR	NR	LR	NR	NR	UR	LR	LR	UR
Nishita 2013[40]	LR	LR	HR	NR	NR	NR	NR	HR	LR	NR	LR	LR	LR	LR	UR
Patja 2012[41]	LR	UR	LR	NR	NR	NR	UR	UR	UR	UR	UR	HR	LR	LR	UR
Pearson 2013[19] (Pearson 2012[42])	UR	UR	UR	NR	NR	UR	NR	NR	UR	UR	NR	HR	UR	LR	HR
Pinto 2015[43]	UR	LR	NR	NR	UR	NR	NR	NR	NR	UR	NR	LR	LR	LR	HR
Ruggiero 2010[44]	UR	UR	LR	NR	NR	NR	NR	NR	LR	NR	NR	UR	LR	LR	UR
Ruggiero 2014[20]	LR	UR	LR	NR	UR	UR	NR	NR	LR	UR	NR	HR	UR	LR	HR

Study[a]	1a	1b	2a	2b	2b	2b	2b	2c	3a	3b	3c	4	5	6	Overall Risk of Bias Rating
Sacco 2009[45]	HR	UR	LR	NR	UR	UR	NR	UR	LR	UR	UR	UR	LR	LR	UR
Safford 2015[46]	LR	UR	LR	NR	NR	NR	NR	HR	LR	NR	HR	HR	LR	LR	HR
Sandroff 2014[47]	LR	LR	HR	NR	HR	NR	NR	NR	HR	HR	NR	HR	LR	LR	HR
Sherwood 2010[48]	UR	UR	UR	NR	NR	NR	NR	NR	UR	UR	NR	UR	UR	UR	HR
Thom 2013 #375[49] (Moskowitz 2103[50])	UR	LR	LR	NR	NR	NR	NR	NR	LR	NR	NR	HR	LR	LR	UR
Turner 2012[51]	LR	LR	LR	HR	NR	NR	HR	NR	LR	UR	NR	LR	UR	LR	LR
Vale 2002[52]	LR	LR	HR	NR	NR	NR	NR	NR	LR	UR	NR	LR	UR	LR	UR
Vale 2003[53]	LR	LR	LR	NR	HR	LR	LR	NR	LR	LR	NR	LR	LR	LR	LR
Van der Wulp 2012[54]	LR	LR	NR	NR	UR	UR	NR	UR	NR	UR	UR	UR	UR	UR	UR
Varney 2014[55]	LR	LR	LR	NR	HR	NR	NR	NR	LR	HR	NR	HR	UR	HR	HR
Wadden 2011[56]	LR	UR	UR	NR	NR	NR	NR	NR	UR	NR	NR	LR	LR	LR	UR
Wayne 2015[57]	LR	LR	LR	NR	NR	NR	NR	NR	LR	NR	NR	HR	LR	UR	UR
Whittemore 2004[58]	UR	UR	LR	NR	HR	HR	NR	NR	LR	HR	NR	UR	LR	LR	HR
Willard-Grace 2015[59] (Thom 2015[60])	LR	LR	LR	HR	NR	NR	NR	NR	LR	HR	NR	UR	LR	LR	UR
Wolever 2010[61]	UR	UR	LR	UR	UR	NR	NR	NR	LR	LR	NR	LR	LR	LR	UR
Young 2014[62]	UR	UR	NR	NR	UR	NR	NR	UR	NR	UR	UR	LR	HR	LR	UR

Abbreviations: UR=Unclear risk, LR=Low risk, HR=High risk, NR=Not reported

[a] The companion paper (noted in parentheses) is not rated separately in this table.

[b] Recall bias

[c] Contamination

APPENDIX D. PEER REVIEW COMMENTS

Question Text	Reviewer Number	Comment	Authors' Response
Are the objectives, scope, and methods for this review clearly described?	1	No - First, thanks so much for taking on this challenging and extremely complicated review. It's clear from the report that a tremendous amount of effort went into conducting the review, completing the huge number of analyses, synthesizing the findings and writing the report! However, after reading the report, I must say that I have serious concerns about the meaningfulness of the findings of the report due to the approaches, and subsequently the methods, used to define health coaching, identify the sample of health coaching studies, and subsequently evaluate the impact of these "health coaching" interventions. As a participant in the planning for this ESP project, I participated in the decision-making about the definitions of health coaching and the choice of inclusion and exclusion criteria, so I must take some responsibility for the subsequent impact of those decisions. However, after seeing how the selection criteria impacted the sample, I am afraid that the criteria used to identify health coaching interventions for the ESP raise serious doubts and limitations about the meaningfulness of the findings. Despite the efforts of the ESP team and key stakeholders to choose meaningful inclusion and exclusion criteria for the ESP review, I believe that the decision to select studies based on authors' "self-identification" of an intervention as a health coaching intervention seriously limits the value of the subsequent analyses and meaningfulness of the findings. Conceptualizations of health coaching vary across investigators, even among those who label their intervention as health coaching. To my point, if I	We thank the reviewer for the careful review of the report, including the critique of current definitions of health coaching and how the decision to use self-identified health coaching interventions may have impacted our results. Any method for identifying literature for complex behavioral intervention/innovation has strengths and limitations. This is even more pronounced when the complex behavioral intervention has not been well defined and there is no consensus on what constitutes key elements of the approach. Health coaching in not immune to these complexities. As illustrated in Wolever's 2011 *Archives of Internal Medicine* commentary, there is currently no agreement on what comprises health coaching. To date, there has also been no research to establish the active ingredients of health coaching intervention. Thus, in close consultation with our key stakeholders and our technical expert panel, we weighed our options for identifying this literature. We jointly decided upon use of self-identified interventions. This approach is supported in the literature; it has been used in at least two other recent systematic reviews of health coaching: (1) Olsen JM, Nesbitt BJ. Health coaching to improve healthy lifestyle behaviors: an integrative review. Am J Health Promot. 2010;25(1):e1-e12, and (2) Kivela K, Elo S, Kyngas H, Kaariainen M. The effects of health coaching on adult patients with chronic diseases: a systematic review. Patient Educ Couns. 2014;97(2):147-157. Also, Wolever's seminal systematic review focused on how health coaching has been defined in the literature: Wolever RQ, Simmons LA, Sforzo GA, et al. A systematic review of the literature on health and wellness coaching: defining a key behavioral intervention in healthcare. Glob Adv Health Med. 2013;2(4):38-57.

Question Text	Reviewer Number	Comment	Authors' Response
		counted correctly, only 14 of the 41 studies in the sample were found to include all 3 of the prioritized components of health coaching identified by stakeholders; only 68% had the highest priority patient-centeredness element; while 5 studies had none of the 3! So, almost 2/3 of the selected studies did not include all 3 high priority health coaching elements. Thus, there is clearly a lack of consistency among the investigators of the selected trials regarding the conceptualization of health coaching. And, because the inclusion of studies was largely based on self-identification of health coaching, it is also highly likely that many interventions that might have been included in the sample, based on the health coaching definition used in the ESP, were not included because they did not use coaching as a descriptor in their title or abstract or key words. To cite just 1 example of the impact of the decision to use self-identification as a key determinant of inclusion, the Pinto et al study that was included in the sample of 41 health coaching studies utilized a physical activity counseling intervention that is almost identical in content and approach to interventions that were utilized by the same investigators in multiple previously published physical activity intervention trials. I was a member of the investigative team for several of these studies, so I am quite sure the intervention approach was the same, except for the use of peers to deliver the counseling intervention in the selected study (I'm happy to share the citations of other trials with you). Yet, because we did not previously use the term coaching in the title, abstract or key words of the previous publications, they did not meet inclusion criteria and were not included this ESP review. I believe there are probably many other examples of investigators who inconsistently labeled their intervention as health coaching. I am also quite sure that there are many studies that applied an	We recognize that any approach to identifying this literature would introduce heterogeneity. Thus, we sought to unpack this complexity by applying a health coaching concordance standard across the identified literature. This concordance score was co-developed with stakeholders, technical expert panel members, and local experts in health coaching. Although we agree that it was surprising to find a high number of studies that did not include all three elements, our intent was not to characterize interventions as meeting health coaching criteria if these three elements were present. Rather, we were interested in examining the relationship between concordance with key elements and select outcomes. We also agree that these key elements are not unique to health coaching. Many behavior change approaches share common elements and, as the reviewer states, there is significant overlap in approaches. It is precisely for that reason that we tasked our collaborators and external experts with prioritizing key elements of health coaching. The reviewer's assessment that these key elements were not applied to a high degree across studies and that there is overlap in approaches are excellent points that we now have stressed to a much higher degree in the study conclusions. Health coaching is an emerging field with shifting definitions across time. Our approach offers a snapshot of the literature at the current time. The heterogeneity of the identified studies underscores the importance of better efforts to distinguish health coaching from other common behavioral interventions. We have expanded our Discussion section to include a broader discussion of the variability of the included study.

Question Text	Reviewer Number	Comment	Authors' Response
		intervention that would meet the definition of health coaching that were not included for this ESP because the intervention was labeled differently (e.g., as behavioral counseling, self-management support, motivational interviewing, health education). Moreover, I would argue that health coaching, as conceptualized in this ESP, is as an interactive "process" that is a core component of a wide range of health behavior change interventions, rather than a specific intervention type. And, the 3 prioritized elements of health coaching specified in the ESP, are key elements of many theoretically derived health behavior change interventions, including those based on self-determination theory, the transtheoretical model, the health belief model, the PRECEDE-PROCEED model, social cognitive theory, motivational interviewing and the 5As (Assess, Advise, Agree, Assist, Arrange). See the Whitlock et al publication for a model that was developed to assist the USPSTF evaluate the impact of health behavior counseling interventions in primary care. The authors also promote the use of the 5As as "a unifying construct to describe behavioral counseling interventions across behaviors". The 5As approach, as described by Whitlock, clearly includes the 3 high priority elements of health coaching specified in the ESP report, along with other key elements that are linked to health behavior change theoretical models. [Citation: Whitlock P, Orleans CT, Pender N, Allan J. Evaluating primary care behavioral counseling interventions: an evidence- based approach, Am J Prev Med 2002;22:267–84.] – Because there is extensive overlap between self-identified health coaching interventions, self-identified health behavior counseling interventions and self-identified self-management support interventions, use of self-identification as a key inclusion criteria	

Question Text	Reviewer Number	Comment	Authors' Response
		elevates use of a descriptive label as THE differentiating feature. And because this label is inconsistently and idiosyncratically applied, it reduces the meaningfulness of the findings of the review. I have offered some other specific examples of the limitations of using self-identified health coaching as a selection criteria in the "additional comments" section below. Unfortunately, at this point in the ESP process, there is not much that can be done to address the inclusion/exclusion issue. However, as authors of the report, it would be helpful to more clearly discuss the limitations of the report, specifying how the limitations raises increased uncertainty about the meaningfulness of the findings. As noted in my response to the question about bias below, I feel that the Clinical Implications and Conclusions sections overstate the meaningfulness of the findings and could more directly reflect the significant methodological limitations. – See also the very recent article by Larsen et al in the February 2017 issue of the J of Behavioral Medicine - "Behavior change interventions: the potential of ontologies for advancing science and practice". The authors offer a very cogent argument for developing and applying common language for characterizing key aspects of behavioral interventions to aid efforts to tease apart elements responsible for impacting behavioral outcomes within specific populations and contexts. Using their "ontology", elements of health coaching would fall into the "Intervention Delivery" class - "Includes mode of delivery including face-to-face, telephone, SMS text, mobile app, website, mass media etc. It also involves style of delivery such as engagement features of an app, or communication style of a counsellor. It also includes duration,	

Question Text	Reviewer Number	Comment	Authors' Response
		amount, and fidelity to designed content". Larsen et al also define Intervention Content - "What is delivered by the intervention in terms of behavior change techniques (BCTs) and intervention functions. BCTs are potentially active ingredients that may be specified in terms of an appropriate taxonomy which may be mapped on to Michie et al's BCT Taxonomy v1 taxonomy". We would have to delve deeper into this ontology to determine where health coaching might fall - my guess is that it would include elements of both "delivery (particularly counselor "sytle") and "content". As Larsen point out, use of a more precise common language will aid evaluation of intervention element impact as well as conditions for effects and mechanisms. As Larsen and colleagues note, "The goal of this ontology is to provide a means of answering the question, 'What works to change what behaviors, for whom, in what situations, how and why?'. This approach may help guide future research on the impacts of key intervention elements, including those identified by the ESP team and stakeholders as prioritized elements of health coaching.	
	2	Yes	
	3	Yes	
	4	Yes	
	5	Yes	
	7	Yes	

Is there any indication of bias in our synthesis of the evidence?	1	Yes - See response to the question on objectives, scope and methods, as well as additional comments. The authors are aware of limitations due to the lack of consensus on a definition of health coaching and the selection of studies based self-identification as health coaching interventions. These are clearly mentioned as limitations on page 79. However, my overall impression upon reading the report is that the authors do not fully appreciate nor specify the potential impacts of the limitations on the results. Though other important limitations are noted (e.g, only 15% of selected studies were judged to have low risk of bias; high levels of heterogeneity; lack of any consistent relationships between study characteristics and outcomes; inability to conduct multivariate analyses of study characteristics), these limitations are not adequately reflected in the language used to summarize findings, particularly in the Clinical Implications and Conclusions section. Given the serious limitations, I respectfully disagree with the conclusions (on page 81) that the "overall results suggest self-identified health coaching have the potential to produce small positive statistically significant effects..." and "compared with usual care, health coaching interventions may be as effective as other behavior change techniques". Though the language, "has the potential to" and "may be as effective as...." reflects uncertainty regarding the strength of evidence supporting the findings, I feel the limitations are such that it is important to be clearer about the limitations and the uncertainty of the findings, especially in the clinical significance and conclusion sections. As noted, there were significant problems reliably characterizing and identifying health coaching interventions an differentiating these from other interventions that focus on health behavior change.	Thank you for these comments. We have expanded our Conclusion section to place greater emphasis on the need for several foundational steps to occur prior to additional research on the effectiveness of health coaching. These proposed steps include the development of consensus definitions of health coaching and the credentials required to become a certified health coach, greater attention to behavior change taxonomies when developing and describing interventions, and more rigorous publication standards requiring more complete description of study design, randomization, and reporting. We have made revisions throughout our Discussion section. We have retained this language in the Conclusion section: “However, the relatively large number of studies at high or unclear ROB and the moderate to high heterogeneity in pooled estimates limit certainty about the interpretation of our findings and the conclusions that may be drawn.”
	2	No	

Question Text	Reviewer Number	Comment	Authors' Response
	3	Yes - It is not biased in the sense of being inconsistent or favoring a specific outcome, but is biased in the sense that evaluation criteria for biomedical studies was applied without adequate attention to the fact that these are behavioral trials that warrant different criteria in assessing quality. Specifically, the rating of unclear bias for behavioral trials that did not blind participants to randomized intervention is not appropriate for behavioral trials of this nature. Hence, the quality of the available research (while still low to moderate and in need of stronger designs) is portrayed as lower than it is. Similarly, the use of wait-lists and usual care (seen as less rigorous for biomedical studies) is more appropriate than most other designs in many of the health coaching studies yet is consistently described in the report as providing "weaker" research. (For example see the work of K. Freedland, editor of Health Psychology. e.g., Freedland KE. Demanding attention: reconsidering the role of attention control groups in behavioral intervention research. Psychosom Med. 2013;75(2):100-102; Freedland KE, Mohr DC, Davidson KW, Schwartz JE. Usual and unusual care: existing practice control groups in randomized controlled trials of behavioral interventions. Psychosom Med. 2011;73(4):323-335.) Most of the health coaching studies are pragmatic effectiveness trials, not efficacy trials, and have the goal of evaluating potential improvement in clinically important outcomes rather than the goal of "analysis in isolation" used in most biomedical studies. I am so sorry that I was not aware of the plans to apply Cochrane criteria to rate quality or I would have raised this issue when my input was first invited. An additional issue related to the lack of understanding of behavioral trials is that weight/BMI is consistently categorized as patient "behaviors" in KQ1 throughout the report. Weight and BMI, though	We thank the reviewer for this important point. We now reference the detailed guidance for the Cochrane Risk of Bias Evaluation Tool in Appendix C (Higgins J, Altman DG. Chapter 8: Assessing risk of bias in included studies. In Cochrane Handbook for Systematic Reviews of Interventions Version 5.0, 2008. Available at: http://handbook.cochrane.org/chapter_8/8_assessing_risk_of_bias_in_included_studies.htm.) These criteria allow for unblinding of participants and an assessment of low risk of bias. Prior to beginning the risk of bias evaluation process, our study team was rigorously trained in how to properly apply these criteria to behavioral interventions. Also we added additional information to our limitations section describing the fact that many studies did not fully report information on their study design for reviewers to properly assess bias. This resulted in many components of bias being deemed "unclear" and overall ratings resulting as "high risk." The reviewer also introduces an interesting idea about the language we use to describe usual care comparators. We have modified our language to map to this excellent point. We have now used the term "weight management" to describe studies assessing the outcome of weight loss as measured by changes in weight (lb/kg) or BMI.

Question Text	Reviewer Number	Comment	Authors' Response
		highly related to behavior, are clinical outcomes - not patient behaviors. (A person can not "do" weight; they eat or don't exercise and weight may be the result, just like dysregulated blood sugar.) That this categorization is inaccurate is also evident when the authors note diagnostic categories of patients, including obesity with the categories of type 2 diabetes, heart disease, etc. (e.g., see p. 12 line 26; and throughout the results - e.g., p. 64 line 23); obesity is a clinical condition not a behavior.	
	4	Yes - In my assessment of current literature, including this study, there is a core misunderstanding or lack of understanding of what health coaching is, and why it is different from case or disease management, education, or directed behavioral instruction. The bias that is pervasive is that of a reductionistic, linear cause and effect model of disease and healing. The greatest potential gains from the introduction of Health Coaching into the current medical system is it bringing an approach that shifts the entire orientation toward the patient/client within our system of health care. Putting Health coaching into a directed intervention like a drug, without understanding its core nature, is doing a disservice to the potential it can bring.	Thank you for these comments. Patient-centeredness was one of our *a priori* key elements of health coaching in the concordance score. In fact, our advisors rated it as the most important element. Thus, a study was given one point for demonstrated use of patient-determined goal or use of self-discovery process and 2 points for patient-centeredness, as this was rated as the main driver of coaching effects by our stakeholders.
	5	No	
	7	No	

Question Text	Reviewer Number	Comment	Authors' Response
Are there any published or unpublished studies that we may have overlooked?	1	Yes - See other comments.	
	2	No	
	3	Yes - In general the search was very thorough, but it was unclear to me why some trials were left out. For example, the Duke study by Edelman et al (2006) was clearly a health coaching trial. Perhaps it was left out because there were multiple elements to the intervention in addition to coaching. However, that was also the case for included trials such as Wennberg (2010), Appel (2011) or Annesi (2011), so selection was unclear to me. It would be helpful to have a clarifying statement regarding how inclusion decisions were made when the intervention had both coaching and other elements (and had a sample with chronic disease).	Our initial literature search captured this Edelman et al study (2006). However, our inclusion criteria states that eligible studies needed to be designed to recruit individuals with one or more chronic medical condition. This study sought to recruit patient with elevated CVD risk, not a pre-existing chronic medical condition. In an assessment of the recruited population, it is possible that up to 40% of the sample may only have had a risk factor and no preexisting chronic medical condition. Further, we consulted with the primary author of this study, who is also an author of this report, and he agrees that the eligibility criteria for this systematic review doesn't map to the population he sought to recruit for his study. We have clarified our eligibility criteria in Table 1 to better reflect populations deemed eligible for this systematic review.
	4	Yes - I believe the original decision to include only RCT's eliminates significant data that can be gained from other published studies,, See Levin, Jeffrey S., et al. "Quantitative methods in research on complementary and alternative medicine: a methodological manifesto." Medical care 35.11 (1997): 1079-1094.	We agree that limiting to RCTs eliminates data that may be contained in other studies with non-RCTs designs. It would have been infeasible to include other designs without limiting the review in other ways, such as limiting populations or outcomes of interest. As our stakeholders were interested in the effectiveness of health coaching across a wide swath of populations and outcomes, and there were ample literature, we limited the evidence base to the most rigorous study design. We, however, now state in our limitations section that restricting to RCT is a limitation.
	5	Yes - I only saw two studies on diabetes prevention program. Health coaching is part of the intervention. Dr. Robert Ackermann has published in this area and I didn't see any of his articles.	Our systematic review's eligibility criteria specify that populations must have a chronic medical condition. Thus, studies of disease prevention would have been excluded. We have clarified our eligibility criteria in Table 1 to better reflect populations deemed eligible for this systematic review.
	7	No	

Additional suggestions or comments can be provided below. If applicable, please indicate the page and line numbers from the draft report.	1	page 2, line 24 - I think you meant to say "decreases" in HbA1c. My comments regarding the Abstract conclusions are noted in my responses to reviewer question on bias above. page 3, line 51 - I have some trouble with the statement, "While health coaching shares common elements with other intervention approaches such as patient education and disease management, health coaching differs in its emphasis on both the overall approach and the process." The first part of the statement is accurate, but the latter part ("health coaching differs in its emphasis....") is an overgeneralization. Though "patient education" may be delivered by a directive or didactic manner by some (e.g, those who have not received training in contemporary health education approaches), health education is, by definition, a process that overlaps quite considerably with the definition of health coaching applied in this ESP project. NCP's Veterans Health Education and Information (VHEI)program defines health education as follows: "Health education is a process that includes any combination of education, information, and other strategies to help Veterans optimize their health and quality of life. Health education programs and services assist Veterans to adopt healthy behaviors, partner with their health care teams, make informed decisions about their health, manage their acute and chronic conditions, and use problem-solving and coping skills." This definition aligns with published health education theory and with models of health education research, training and practice that feature collaborative, patient-centered approaches. To cite just one example, in 1980, in their book, Health Education Planning: A Diagnostic Approach, Green, Kreuter and others described the PRECEDE Model, which featured assessment of individual predisposing, enabling and reinforcing factors when developing and delivering health education interventions. The PRECEDE approach requires	The direction of effects is reflected in the point estimates. There are numerous definitions of health education that range from a narrow intervention approach to a field of social science. However, health coaching is conceptualized as an intervention approach which can be considered one tool, or approach, used in the field of health education. Our intent in the introduction was not to draw a contrast between the field of health education and the intervention approach of health coaching. We have modified the introduction to better emphasize the differences in approaches between traditional patient education, not the field of health education, and emerging models of collaborative intervention, like health coaching. The reviewer asks how we defined “patient education only” and “disease management only.” Again, we allowed the study authors to categorize their own intervention approaches. The moderators we explored were suggested by, and developed with, our stakeholders. Training, supervision and in-study monitoring of fidelity were not selected as key variables to explore. We agree with the reviewer that these are important component of implementing high quality behavioral interventions. It is important to note that these elements are not routinely reported in outcomes papers of behavioral interventions, especially in-study monitoring of treatment fidelity. Even when fidelity is reported, the description usually contains the process of assessing treatment fidelity and not the degree to which the implementation of the intervention maps to the intended content and approach of the intervention. Thus, conducting such analyses are likely infeasible. We did routinely collect how interventionists were trained, if these descriptions were present in the studies. We have added information on the variability in coach training in the description of the included studies. Also our future research table lists training as a key area for future research.

		actively engaging patients, exploring their values, needs and preferences, and tailoring interventions to these elements. Health education interventions have also been informed by other theories and models (e.g., Health Belief Model, Social Cognitive Theory, TTM, Self-Determination Theory) that emphasize an interactive patient-centered process. See my next comment as well. page 3, line 55 - 58. The statement, "traditional health education interventions are more likely to be expert directed, task oriented and focused on disease-specific content, whearas health coaching is collaborative, client-centered and more likely to be focused on the whole person", is misleading and and does not accurately characterize the current state of health education programs, most of which embrace collaborative patient-centered principles. Please see my previous comment regarding health education as a collaborative patient-centered process. Another example of the collaborative nature of health education programming is Holman and Lorig's characterizations of self-management and self-management support. (See: "Patients as Partners in Managing Chronic Disease, BMJ, 2000). Indeed, contemporary health education programs actually feature "health coaching" components (as defined in the ESP) and refer to them as critical and essential elements. (Note: the VHA VHEI Program has published a "Veteran-Centered Health Education Workbook" that features strategies for enhancing the patient-centeredness of VHEI programming.(It is available upon request). Though I understand the ESP investigators' need to focus in on health coaching interventions and could be differentiated from large volume of health education and self-management support interventions that have been investigated in clinical trials, it is not accurate to imply that all patient education interventions are any less collaborative than health coaching interventions. The presence of	We have added to our Discussion section that there is overlap in the concordance elements and elements of other health behavior change approaches. It is an interesting idea to compare the proportion of low risk of bias studies in this report to systematic reviews of other interventions designed to impact chronic illness outcomes. There is a high degree of variability in how study quality is assessed (tool used, domains assessed) and reported (overall score qualitative score, numerical score, by bias item vs overall score) across systematic reviews. Also these risk of bias assessment approaches and applications vary based on the type of interventions being reviewed. Further, while PRISMA calls for an assessment of risk of bias, many systematic reviews do not conduct such assessments. Thus, comparing our assessment of overall risk of bias to other studies is likely infeasible and unadvisable based on the reasons stated here. In the Clinical Implications section we state that “For HbA1c, there is consensus that improvements of 0.3%, the summary effect found in this study, are clinically relevant changes. . .” While we have added that other nonpharmacologic interventions have produced equal or greater effects, it should be noted that the Chodosh et al (Ann Intern Med 2005) study included a definition of disease self-management that study authors stated was very broad and likely included studies that other may not have included in a review of disease self-management, summary effects demonstrated important heterogeneity, which means that this effect size must be interpreted with caution, and the review is over 10 years old making comparisons to this literature review imprudent. Regarding the Pinto et al study, we have stated throughout the report the strengths and limitations of our approach to identifying the literature on health coaching. We are in agreement that several of the interventions included in the weight management sections of the report do not meet criteria set forth for comprehensive lifestyle

		health coaching consistent elements in patient education programs may help explain the ESP's finding of a limited effect of health coaching vs "active comparators", as many of the "active" health education comparators also have patient-centered, collaborative "whole person" components. On the other hand, exclusion of patient education interventions from the sample of health coaching interventions may have limited the likelihood of finding positive effects for health coaching on the chosen outcomes. For example, self-management support interventions for diabetes, which I have argued almost always include process elements that overlap substantially with the elements of health coaching, have demonstrated a positive effect on HbA1c in previous meta analyses (See Chodosh, AnnIM 2005). The exclusion of these "health-coaching like" patient education interventions also limited the analysis of processes that contribute to positive intervention effects. page 5, Table 1. How was "patient education only" and "disease management only" defined? How confident were the reviewers that those identified as patient education and disease management only did not meet the criteria for inclusion as a health coaching intervention? It seems like many self-management support interventions excluded. I noted the list of included studies does not include most Lorig et al self-management support program (CDSMP) interventions and Heisler's peer coaching for diabetes self-management studies, both of which had robust "health coaching" elements. page 8 - The analyses did not include training, supervision and in-study monitoring of fidelity as a potential source of variability. I realize that the number of possible determinants of variability had to be limited for this ESP. A more targeted analysis that includes these potential fidelity-related sources of heterogeneity should be considered in subsequent	interventions. It was not the charge of this report to assess CLI and, thus, beyond the scope of these analyses. We have stated that many of the outcomes used in our analyses were not the main outcomes of each of the studies. Pooling secondary outcomes is not an uncommon practice in systematic review science. Thank you for your ideas on future research. We have expanded our list of future research, including a call for some formative research on the key elements of health coaching that distinguish it from other behavioral approaches. We have modified our concluding statement about the effectiveness of health coaching in relationship to other behavioral approaches and have reiterated that there were multiple adjunctive intervention supports across the included trials in response to the reviewer's keen comments.

		evidence synthesis reviews of health coaching or health behavior counseling interventions. page 8 - 9 - rating of prioritized elements of health coaching. This is a strength of the analyses. Again, I will point out the 3 elements that were chosen are common to most "modern" (and particularly 21st century) health education and self-management support interventions. Even when the SMS is offered in the setting of a chronic condition that focuses on a specific set of self-management behaviors (e.g., the American Association of Diabetes Educators "7"), self-management support interventions feature patient-centered approaches, offer a large menu of possible diabetes SM goals, and engage the patient in an active discovery process.(Note - see Fisher, Ecological Approaches to Self-Management, AJPH, 2005 and the Resources and Supports for Self-Management (RSSM) measure that was developed to capture the elements of effective SMS - https://www.ncbi.nlm.nih.gov/pubmed/18669813) page 12, line 25 - "Most studies recruited populations with type 2 diabetes (n = 18) I think you meant to state, "the most common population recruited was type 2 diabetes". page 12, line 44. "only 15% trials (6 studies) had a grade of low risk of bias.....34% of trials (n = 14) had a high risk of bias". It would be useful to comment either here or subsequently on the very small percentage of studies that were judged as low risk of bias. How does this compare with other similar meta analyses of interventions designed to impact chronic illness outcomes? page 22 - Studies reporting change in HbA1c. Note that the mean difference of .3% found for self-identified health coaching vs an inactive comparator is well below the .81% difference found in the meta-	

		analysis of self-management support interventions among older adults with diabetes conducted by Chodosh et al, AnnIM, 2005). page 26 - Effects on Physical Activity. See my comment in response to the methods question regarding Pinto et al PA studies. page 29 - Effects on Weight, BMI. As previously noted, the selection criteria focused on self-identification of health coaching and this has led to inconsistent inclusion and exclusion of weight loss intervention trials. For example, 2 of the 3 NIH-supported POWER trials are included in the ESP review (Wadden and Appel are included, Bennett et al., ArchIM, 2012 is not), even though the interventions for all 3 trials based their obesity interventions on the Diabetes Prevention Program trial, a comprehensive lifestyle intervention, and was adapted to be compatible for delivery within primary care settings. (Note: Systematic reviews and meta-analyses of behavioral weight loss interventions, conducted by AHRQ, AHA/ACC/TOS, as well as the VA/DoD Clinical Practice Guideline, identified core intervention elements associated with clinically significant weight loss. These reviews also concluded that interventions that don't meet the criteria for CLI are less likely to produce clinically significant weight loss. Based on these systematic reviews of evidence , the AHA/ACC/TOS and VA/DoD guidelnes recommend the provision of comprehensive lifestyle interventions (CLI) as a core element of overweight and obesity management. The VA/DoD CPG specifies that CLIs must include 3 key components (dietary, physical activity and behavior change components), and at least 12 clinical sessions in 12 months.) Several of the interventions included in the ESP analysis do not meet criteria for a CLI. Moreover, many included a weight loss outcome measure, but did not focus on weight loss as a primary objective. (This was mentioned but not	

		emphasized by the authors). Moreover, as discussed by the authors, the non-health coaching elements of the interventions that were included in the ESP's BMI change and weight loss analyses are quite heterogeneous. As a result, the meaning of the findings of the ESP analyses on BMI and weight are uncertain and potentially confusing to readers. The small effects found for reducing BMI across highly heterogeneous studies, populations, and contexts does not, in my opinion, provide support for health coaching as an independent contributor to weight loss and once certainly shouldn't imply from this finding that health coaching may be as effective as a CLI (yet that is what the conclusion implies). Note also, that, in the ASPIRE VHA-based trial, cited on page 32, all 3 interventions met the criteria for a CLI and all produced significant weight loss. Moreover, at follow-up, there was no significant difference in weight loss across arms, despite the increased support provided to interventionists who delivered the enhanced health coaching arm. From the perspective of NCP, it is critically important that VHA continues to focus on disseminating and implementing CLIs for weight management that meet the criteria recommended by current VA/DoD guidelines, which in turn are based on rigorous synthesis of the best available evidence. On the other hand, gaps in our understanding remain about how to best help patients to engage and participate in weight management interventions as well maintenance activities when they are successful. Focusing on the 3 priority elements of health coaching, as a COMPONENT of an evidence-based weight management intervention may be helpful in that regard. That might be a very fruitful area for future research - which might be one recommendation stemming from this report. However, the final version of this report should at	

Question Text	Reviewer Number	Comment	Authors' Response
		least mention the current CPG recommendations for weight management, including the recommendation for offering CLI. (Note also 2 publications from Wadden, one a systematic review of behavioral treatment of obesity for patients seen in primary care, and a quite recent review article from NEJM. Both reinforce the importance of offering CLI, while also considering ways to make weight management interventions more accessible and impactful, particularly in primary care settings. (Wadden et al. Behavioral treatment of obesity in patients encountered in primary care settings: a systematic review. JAMA. 2014 Nov 5;312(17):1779-91. doi: 10.1001/jama.2014.14173; Heymsfield SB, Wadden TA.; Mechanisms, Pathophysiology, and Management of Obesity. N Engl J Med. 2017 Jan 19;376(3):254-266. doi: 10.1056/NEJMra1514009.))	
	2	The most significant issue in assessing the effectiveness of health coaching is that health coaching is an emerging profession, and is therefore not well or consistently defined. While this is stated in the review, it is not given the emphasis it deserves. While the review states that one of the characteristics assessed was the training of the 'self identified coaches' (page 6), this is not explored in the article. There is one statement that in all of the studies included there is only ONE trial that used a certified health coach (pg 12)- that is a SIGNIFICANT finding. A very critical issue is that when people self identify as a health coach, at a time when the profession has not been defined, there is huge disparity in the intervention and therefore "mixed results." As I read the descriptions, it seems to me that the common denominator is more likely to be motivational interviewing, and in some studies, even case management, but not what the profession is now defining as health coaching.	We agree with this reviewer and have further highlighted that only one study describes use of "certified health coaches" and the impact that may have on heterogeneity of treatment effects.

Question Text	Reviewer Number	Comment	Authors' Response
		This in no way diminishes the results of the systematic review, but I do feel this factor is not appropriately emphasized. To use an illustration from my profession, it is parallel to trying to draw conclusions about labor and delivery outcomes when looking at pregnancy outcomes of women managed by ObGyns, nurse midwives, and lay midwives combined - all trained in different approaches with vastly different levels of training, and yet all included in the analysis.	
	3	I applaud the authors for this enormous project that entailed countless hours of careful attention. Thank you for this work! I understand why the authors focused on interventions self-described as health coaching given the nascent state of clear definitions for the approach, and yet I am disappointed that we are left with a confounded presentation of the findings of health coaching interventions. Many of the included trials do not fit with the emerging definitions of health coaching [e.g., as put forth in the literature by Wolever, Simmons et al, 2013 (and adopted by the National Health Service in the UK, or that put forth by Olsen & Nesbitt, 2010 or by the International Consortium for Health & Wellness Coaching in partnership with the National Board of Medical Examiners]. The fact that the interventions studied include many that are not health coaching is obvious in the low percentages of trials that met any of the three critical elements defined by stakeholders. That said, it is very useful to have such a comprehensive and systematic work that does show the state of the literature, with all of its problems. My biggest concerns regard the treatment of this literature as if it were biomedical in nature rather than behavioral (as already noted above). Aside from that, I offer specific comments to consider, some of them just typos to correct and some that at are more important conceptually.	Thank you for these observations and your careful reading of the report. We have corrected the typos on the text and provided further clarification in the places the reviewer identified. Moreover, we have clarified our use of risk of bias elements above as they apply to behavioral interventions. We have clarified that the duration described in the study characteristics pertains to the active intervention phase and not the outcome assessment window. Unless otherwise noted, effect estimates are means at follow-up. We have clarified this in our methods section. For weight outcomes and A1c we preserved the natural units and used mean differences (MD) and for other outcomes that were more variable, we used standardized mean differences (SMD). This is noted on each forest plot and in the text of each results section per outcome. We considered including patient activation as an outcome of interest in this study as it is an emerging outcome in health coaching. As it is an emerging outcome of health coaching, our stakeholders thought that it would not be reported consistently, especially in the early literature, prior to the availability of a measure of patient activation. Thus, per the guidance of our technical experts, we collected and analyzed self-efficacy. We have added this idea as a future research topic.

Question Text	Reviewer Number	Comment	Authors' Response
		p. 1 line 57, remove "use" in "use used" p. 2 line 24, "increases" should read "decreases in HbA1c" p. 3 lines 14-15 unclear: does this mean 828 hrs per yr to treat all patients that had any of the top 10 chronic diseases? p. 3 line 42: coaches do not provide advice, including motivational advice. Better to say "is to use motivational processes, ..." p. 3 line 51 "may not be trained therapists" - in fact, the vast majority are not, so better to say "but only the minority are trained..." p. 4 lines 11-12: the report was commissioned for these reasons, and I know why it understandably shifted a bit. The reader won't know this however, so an additional sentence somewhere might make it more clear why quality of life and pt satisfaction were not evaluated p. 4 line 37 - as noted above, weight is not a behavior p. 4 line 50 - there are several places where "certified" health coaches are mentioned. (Also seen on p. 12 line 12, p. 35 line 39, p. 69 line 5.) The naive reader (and perhaps the authors) might assume that "certified" in this context means something in terms of skill level, when it does not. In fact, at this point in time, there is no single certification that implies additional skill behind its brand identity. You can literally go online, pay a fee and become certified without the demonstration of a single skill! The International Consortium of Health and Wellness Coaching is working to change this, but thus far, it is	The p-values in the forest plots pertain to the test for heterogeneity statistic, I^2, and not the summary effect estimates. In reference to the use of the quality rating tool used in this study, review investigators were taught how to apply these criteria to behavioral studies, including when blinding of participants was not feasible.

Question Text	Reviewer Number	Comment	Authors' Response
		misleading to note "certified" as a way to distinguish the coaches in one study versus another (even if those authors did). I strongly recommend removing that distinction. p. 6 line 14 - weight is not a patient behavior but a clinical outcome p. 7 line 13 - it is inappropriate to assume that blinding participants in an active behavioral trial is an indicator of quality research. Please see discussion and refs above. p. 7 line 55 - only six had more than one arm?? (UC vs coaching is two arms) Do you mean more than one active arm/intervention beyond usual care? p. 8 line 25 - were funnel plots explored? p. 8 line 54 - The technical expert from Vanderbilt is from the Vanderbilt University Medical Center (there is no school of PMR, but only a Department of PMR, and of Psychiatry) p. 9 lines 26-27 - Did you characterize an intervention as using pt-determined goals if the participants had any input at all? Many trials have both expert/professional input as well as participant input. For example, one might get an exercise prescription (expert driven goal) but fine-tune it for their readiness. Or get the goal from the expert but have exercised some pt-determination in signing up for a study on a given behavior (e.g., aerobic exercise) in the first place. p. 12 lines 43-44 - An inclusion requirement was that there was at least 6 month outcomes data, so how could only 80% of the trials last months or longer?	

Question Text	Reviewer Number	Comment	Authors' Response
		p. 12, line 56 - weight is not a behavior p. 16 line 25 - Luley study comparator column is unfinished: "then control group left??" p. 16, line 49 - what was the intervention duration? max of 3 X per week for how long? p. 17 line 26 - needs bullet in third column p. 21 line 22 - replace "medical assistant in psychology coach" with "psychotherapist" or with "masters level psychologist" - there were no medical assistants in this trial p. 21 line 48 - Somewhere, (and most likely in the discussion, but it was apparent to me here) it would be useful to note that one of the challenges with health coaching trials is that since goals are self-determined, participants work on different things and inclusion criteria is not targeted to a single variable. It is easier to get significance for change in A1c, for example, when you only include those with elevated A1c to start with. But by glancing over the starting A1c values of the various trials, it is obvious that elevated A1c was not an inclusion criteria for many studies. This is similar for other conditions; when you don't start with elevated weight with all participants, harder to see a change in weight, etc. Hence, trials designs need to consider this in the future. p. 24 Figure 2 - It is not clear what the mean (SD) values represent. It appears that they are sometimes the post-tx mean (without consideration for baseline) and sometimes the mean change for a given group. Is there a way to clarify this, or perhaps at least spell out that these differences are neutralized in the meta-analyses? This is similar for the other Figures - e.g., Figure 3 - steps or minutes?	

Question Text	Reviewer Number	Comment	Authors' Response
		p. 24 line 58 - comma needed after cardiovascular disease p. 32 line 31 - replace "students" with "participants" p. 34 Figure 6 - spacing and size could be better p. 35 Figure 7 - same p. 35 line 38 - remove "study" after ROB p. 36 line 47 - drop the "ed" on addressed p. 37 line 35 - I wonder why you left out trials that measured pt activation as this concept is highly related to self-efficacy? In fact, it is defined as believing you have the confidence, skills and behavior to manage your health condition while self-efficacy is believing you will succeed in a task, goal. Knowing the stellar reputation of your group, I'm sure you considered this; might be nice to have a sentence about the decision to omit such as patient activation is a somewhat "hot" topic. p. 38, line 18 - colon after wait-list should be a comma p. 41 line 30 - reported should be reports or reporting p. 41 - My significant concerns about measurement of performance bias are noted above p. 42, line 5 - The fact that about half of the trials did not report complete outcome data and did not include attrition data is a much better indicator of the low quality of the behavioral trials than the blinding issue. I'd state this instead.	

Question Text	Reviewer Number	Comment	Authors' Response
		p. 43 - A legend would be nice for Figure 10 p. 44 line 31 - add "of" before health coaching p. 45 line 12 - add "s" to finding p. 50. lines 8, 25, and 34 - I may not be reading the table correctly, but I thought that the p values in the bolded Summary (lines 25 and 24) were the overall p value for the pooled estimate. If that is true, then there is an error here: line 8 notes that only the pooled estimate for score of 4 was significant but the p values suggested that score of 4 was not and the estimate for score of 3 was. p. 51 line 9 - "varaibility" has a typo p. 53 lines 15-16 - seems strange to me to group together peers and trained educators p. 53 line 29 - should read with "a score of 1" rather than scores p. 54 line 8 - remove "and" p. 54 line 22 - do you mean video rather than video chat (as in chat thru texts)? p. 56 line 45 - should read "effect of intervention does on change in physical activity" - not on health coaching p. 58 line 32 - monthly coaching for how long? p. 58 lines 43-44 - Again, I'm now doubting that I'm reading the figures right, but if the p values in the bolded Summary (lines 13 and 19 on Figure 22, p. 61) were the overall p value for the pooled estimates, then something is wrong. The text (p. 58 lines 43-44)	

Question Text	Reviewer Number	Comment	Authors' Response
		indicates that the first pooled estimate is sig, but the p noted is 0.51) p. 59 line 41 - Same problem - the text on p. 59 line 41 does not agree with the p value in Figure 21, line 14. Is it significant or not? Looks like it is. p. 64 line 15 - should it say "synthesis FOR this outcome"? p. 65 line 3 - Figure title would more accurately read "Effect of HC on CHANGE IN BMI..." p. 65 line 54 - add "participants" after CVD p. 66 line 12 - change "of" to "on" as in effects on BMI p. 66 line 14 - change "or" to "on" in the figure title p. 67 line 19 - Again, the text and figure don't match. Line 19 suggests that the in-person pooled estimate is not significant but Figure 28 notes p of 0.014 (unless you are adjusting the P, which is not stated anywhere) p. 67 line 24 - Figure title would be more accurate if it read "...on CHANGE IN BMI..." p. 69 line 16 - please make this read "produced ONE OF the largest point estimateS" bc it's not the largest; Varney (2014) and Ma (2013) were both larger. p. 70 lines 38 and 40 - what do the numbers (98), (519 and 626) mean? p. 70 line 46 - Up to this point in the report, behavioral health providers have not been considered/categorized as healthcare providers.	

Question Text	Reviewer Number	Comment	Authors' Response
		Although I disagree with that categorization, it would be better to use language here that is consistent. p. 71 line 24 - drop "groups of" so that it reads "other 2 studies" p. 73 line 20 - The p value of 0.28 does not suggest a trend toward significance as stated on p. 72 line 39. p. 73, lines 38 -39 - Similarly, the text does not agree with the p values on Figure 34. P of 0.94 is not significant (again, if I'm reading the figures correctly) but 0.038 does trend that way. p. 74 line 35 add "s" to element p. 75 line 14 - is this p value correct? Doesn't look that way. p. 75, line 43 - weight loss is not a behavior p. 75, line 57 - As noted above, coaches do not give advice, including motivational advice p. 76 line 6 - add "s" to patient p. 76 line 11 - add "an" before a priori p. 76 line 17 - I'd re-categorize weight/BMI as a clinical outcome p. 76 line 29 - The fact that only 68% were even patient-centered tells you that many were not health coaching interventions. p. 77 line 51 - change "inconsistence" to "inconsistency" p. 78 line 29 - add "on" after "based"	

Question Text	Reviewer Number	Comment	Authors' Response
		p. 78 line 35 - conducting needs a "t" p. 79 line 4 - add "to" after "amenable" p. 79 line 19 - at what time point did the 665 steps/day arise? Is it possible to estimate, average or say in interventions of at least X week duration? p. 79 line 27 - add "per day" after kcal p. 79 line 45 - I'm happy to see this conclusion that HC is likely as effective as other self-management approaches" and it may be useful to add a further sentence that underlines that further work is needed, not only about how it is distinct from other behavioral, patient-focused approaches but that further work is needed to determine when i may be most appropriate, and that clarification studies are needed to determine appropriate background and coach training needed. p. 79 line 46 - chance "in" to "is" p. 79 line 53 - I disagree with that there was "careful quality assessment" - it was indeed very careful and thorough but for biomedical rather than behavioral designs. p. 80 lines 4-6 - It would be nice to highlight that the intervention diversity, although you valiantly attempted to unpack it, it highly problematic in interpreting the results. This seems underplayed to me. p. 80 line 19 - The report doesn't really include number of providers, perhaps you meant number of contacts?	

Question Text	Reviewer Number	Comment	Authors' Response
		p. 80, line 27 - Drop "certified" as the distinction is meaningless at this point in time. p. 81 line 15 - I'd remove "health coaching-related content" as this implies that the investigators don't really understand what coaching is, and assume it is a content-driven intervention p. 81 - Twice in the chart, "non-randomized controlled before-and-after studies" Is noted. Is that not the same as prospective trials (which is also mentioned)? p. 81 line 58 - I'd add that additional research is needed regarding the very definition of coaching. p. 81 - A couple important implications that are not mentioned should be considered. First, the time-course of health coaching trials is quite tricky; people must practice behavior change long enough to adequately demonstrate the subsequent shift in biological measures; short trials (most often funded) make it hard to evaluate longevity and impact of behavior change. Second, as I mentioned above, the fact that participants self-determine their goals means that people are often working on different behaviors and have different outcome goals. Hence, it is challenging to create inclusion criteria that maximizes the capture of potential change. Finally, the conclusions would be another great place to mention that use of biomedical quality ratings for behavioral trials is usually inappropriate, and that UC or wait-lists are often the best controls for behavioral effectiveness trials. In addition, I don't imagine that an ESP would be willing to go here, but the truth is that the focus on RCTs themselves may limit the evaluation of effectiveness when the intervention is driven by patient-preference. Happy to provide refs for this if you'd entertain a comment on such.	

Question Text	Reviewer Number	Comment	Authors' Response
		p. 107 line 37 - I wish you had NA as an option for behavioral trials where it is literally impossible to blind the participants to which condition they are receiving p. 108 - For detection bias, how did you handle self-report instruments in terms of blinding? If the assessment personnel was blinded, but the participants appropriately were not, these could not be blinded either. p. 110 format issue betweens lines 4 and 7 Despite my difference in opinions and my many comments, I am grateful for your hard work, and appreciate what you have contributed to this emerging field.	
	4	The greatest Issue I had with the report was the lack of application of the clear description of the intervention that is then required to be applied in the studies. The intervention of Health Coaching was defined, yet included studies didn't have to actually about that predefined intervention. Yet, the study drew conclusions about the effectiveness of Health Coaching, when Health Coaching as it was defined is not even the confirmed intervention. For example, some included studies had NONE of the 3 defined key elements of Health Coaching. An effective definition of an intervention like Health Coaching needs to be clear about not only what it is, but what it isn't. It isn't directive or prescriptive. It isn't predetermined in frequency or agenda by the providers or study administrators. This design implies a deep seated lack of understanding of the principles of Health and Wellness Coaching, besides the 3 determinants discussed, which appeared to be optional. See Wolever, Ruth Q., et al. "A systematic review of the literature on health and wellness coaching: defining a key behavioral intervention in	We thank the reviewer for these thoughtful comments. As discussed above, any method for identifying literature for complex behavioral interventions has strengths and limitations. This is even more pronounced when the complex behavioral intervention has not been well defined and there is no consensus on what constitutes key elements of the approach. Such is the case for health coaching. Thus, in close consultation with our key stakeholders and our technical expert panel, we weighed our options for identifying this literature and jointly decided upon use of self-identified interventions. This approach is supported in the literature, including Wolever review: (1) Olsen JM, Nesbitt BJ. Health coaching to improve healthy lifestyle behaviors: an integrative review. Am J Health Promot. 2010;25(1):e1-e12 (2) Kivela K, Elo S, Kyngas H, Kaariainen M. The effects of health coaching on adult patients with chronic diseases: a systematic review. Patient Educ Couns. 2014;97(2):147-157.

Question Text	Reviewer Number	Comment	Authors' Response
		healthcare." Global Advances in Health and Medicine 2.4 (2013): 38-57. There is great potential damage by making conclusions about whether Health Coaching is effective or not, when it appears that the intervention evaluated didn't even meet criteria to be considered health coaching. The International Consortium Health and Wellness Coaches (http://ichwc.org) under the scope of determining eligibility for sitting the National Board exam, requires that coaching sessions on practice logs must be a minimum of 3/4 specifically coaching, and NOT education or instruction. While there was elegant statistical analysis done, I feel decisions made early about what to include in the review did not lead to advancing the understanding of the field or fair assessment of the effectiveness of such heterogenous interventions.	(3) Wolever RQ, Simmons LA, Sforzo GA, et al. A systematic review of the literature on health and wellness coaching: defining a key behavioral intervention in healthcare. Glob Adv Health Med. 2013;2(4):38-57. Our intent in creating the concordance score was not to characterize interventions as meeting health coaching criteria if these three elements were present. Rather, we were interested in examining the relationship between concordance with key elements and select outcomes. Thus, a study could contain none of the prioritized elements and still be included. We have expanded our limitations section to more fully capture the heterogeneity of the included studies and the impact this has on our findings.
	5	Thank you for the opportunity to review this research synthesis report. I was very impressed with the rigor and breadth of the report. I believe it has excellent insights for clinical practice (despite the young state of health coaching literature). The area of the report that was lacking (in my opinion) was more robust explanation and strategy around "Concordance." Concordance (which is associated with intervention fidelity) was loosely and subjectively defined. Three elements were chosen based on the Wolever (2013) article: Patient-centeredness, patient-determined goals, and use of self-discovery. A score of 0 - 4 was assigned, with one element receiving a maximum of two points. The rationale for this was that "(patient-centeredness) was rated as the main driver of coaching effects by our stakeholders." Given the quantitative rigor of this report, this particular decision seems quite subjective. In my clinical experience, patient-centeredness is a pre-requisite of an effective	Thank you for these comments. We developed the list of key elements and the subsequent concordance score as an exploratory approach to unpacking variability in treatment effects. As described above, these elements were not used as part of the eligibility criteria. We have expanded our limitations section to add more emphasis to the high level of intervention variability in this review. Reviewer disagreements about presence or absence of key elements were arbitrated in the same way as all other data abstraction elements as described in the methods section. The two reviewers worked to come to consensus. If they could not, then a third reviewer, most often the PI, broke the tie. We, unfortunately, did not track inter-rater reliability between reviewer pairs on these elements.

Question Text	Reviewer Number	Comment	Authors' Response
		coaching experience, but it is not the most heavily weighted. In fact, the Motivational Interviewing literature shows that counselors/coaches who are effective at relationship building but not effective at focusing sessions on patient-centered goals have less impact on treatment outcomes. Additionally, two reviewers rated the papers for concordance but their inter-rater reliability was not reported. Where they did differ in their ratings, there was no defined process for how differences were reconciled. I strongly recommend that you include more information about your treatment of Concordance in your methodology and in your conclusions and limitations sections. My reason for emphasizing this point is this: If the intervention delivered did not adhere to treatment fidelity, you cannot explain outcomes. If something failed to have impact, it may be because that "something" was not accurately delivered. Likewise, if something had impact but not treatment fidelity, you cannot say how the outcome was achieved. If Concordance was more completely and defensibly described, I would have given this review an "excellent" rating. A few additional comments are included in my uploaded text. Thank you again for the opportunity to review this paper.	
	7	Overall it is a nice review that provides some interesting data, which have the potential to move the field forward. The biggest concerns I have with the review are twofold. First, there appears to be an underlying assumption that health care providers somehow have training that makes them experts in coaching. Having trained nearly 400 health care providers in coaching, I can veritably say, this is not true. For example, on page 76, the authors write, "all	Thank you for these thoughtful comments and your careful review. We agree that training as a health care provider is not equivalent to training as a health coach. We did not mean to imply so in this report. We systematically captured they type of person delivering the intervention and labeled this as "coach type" but the reviewer is correct that this should be corrected so as not to confound interventionist's discipline (eg, medicine, nursing) with coach training. We

Question Text	Reviewer Number	Comment	Authors' Response
		delivered by the following: certified health coach, study-trained coach, nurse, and a coach with an unspecified training or discipline." Because one is a nurse or a physician does not make one a coach. This needs to be clarified throughout the text. Relatedly, in the limitations section it needs to be made clear that there really are no examples where trained health professionals who are also certified coaches provide the interventions. This is a significant gap in the literature, and based on the data presented in this review, it appears that more studies need to examine the efficacy of coaching interventions where health care providers trained in health coaching provide the coaching. It would be nice to see this in the implications - what would it look like to move forward effectively in the realm of "innovative and rigorous study designs to explore the central elements that distinguish health coaching from other behavioral counseling and self-management approaches and how these unique elements impact clinical and behavioral outcomes?"	have modified language throughout the report to reflect this distinction.