A systematic review of empirical studies examining mechanisms of implementation in health

Cara C. Lewis 1,2,3*, Meredith R. Boyd 4, Callie Walsh-Bailey 1,5, Aaron R. Lyon 3, Rinad Beidas 6, Brian Mittman 7, Gregory A. Aarons 8, Bryan J. Weiner 9 and David A. Chambers 10

Abstract

Background: Understanding the mechanisms of implementation strategies (i.e., the processes by which strategies produce desired effects) is important for research to understand why a strategy did or did not achieve its intended effect, and it is important for practice to ensure strategies are designed and selected to directly target determinants or barriers. This study is a systematic review to characterize how mechanisms are conceptualized and measured, how they are studied and evaluated, and how much evidence exists for specific mechanisms.

Methods: We systematically searched PubMed and CINAHL Plus for implementation studies published between January 1990 and August 2018 that included the terms “mechanism,” “mediator,” or “moderator.” Two authors independently reviewed title and abstracts and then full texts for fit with our inclusion criteria of empirical studies of implementation in health care contexts. Authors extracted data regarding general study information, methods, results, and study design and mechanisms-specific information. Authors used the Mixed Methods Appraisal Tool to assess study quality.

Results: Search strategies produced 2277 articles, of which 183 were included for full text review. From these we included for data extraction 39 articles plus an additional seven articles were hand-entered from only other review of implementation mechanisms (total = 46 included articles). Most included studies employed quantitative methods (73.9%), while 10.9% were qualitative and 15.2% were mixed methods. Nine unique versions of models testing mechanisms emerged. Fifty-three percent of the studies met half or fewer of the quality indicators. The majority of studies (84.8%) only met three or fewer of the seven criteria stipulated for establishing mechanisms.

Conclusions: Researchers have undertaken a multitude of approaches to pursue mechanistic implementation research, but our review revealed substantive conceptual, methodological, and measurement issues that must be addressed in order to advance this critical research agenda. To move the field forward, there is need for greater precision to achieve conceptual clarity, attempts to generate testable hypotheses about how and why variables are related, and use of concrete behavioral indicators of proximal outcomes in the case of quantitative research and more directed inquiry in the case of qualitative research.

Keywords: Mechanism, Moderator, Mediator, Determinant, Implementation, Causal model, Theory
Contributions to the literature statement

- This is the first systematic review of implementation mechanisms across health that assesses the quality of studies and the extent to which they offer evidence in support of establishing mechanisms of implementation.
- We summarize nine examples of models for evaluating mechanisms.
- We offer conceptual, theoretical, and methodological guidance for the field to contribute to the study of implementation mechanisms.

Background

Implementation research is the scientific evaluation of strategies or methods used to support the integration of evidence-based practices or programs (EBPs) into healthcare settings to enhance the quality and effectiveness of services [1]. There is mounting evidence that multifaceted or blended implementation strategies are necessary (i.e., a discrete strategy is insufficient) [2, 3], but we have a poor understanding of how and why these strategies work. Mechanistic research in implementation science is in an early phase of development. As of 2016, there were only nine studies included in one systematic review of implementation mediators specifically to the field of mental health. Mediators are an intervening variable that may statistically account for the relation between an implementation strategy and outcome. We define the term mechanism as a process or event through which an implementation strategy operates to affect one or more implementation outcomes (see Table 1 for key terms and definitions used throughout this manuscript). Mechanisms offer causal pathways explaining how strategies operate to achieve desired outcomes, like changes in care delivery. Some researchers conflate moderators, mediators, and mechanisms [6], using the terms interchangeably [7]. Mediators and moderators can point toward mechanisms, but they are not all mechanisms as they typically are insufficient to explain exactly how change came about.

In addition to these linguistic inconsistencies and lack of conceptual clarity, there is little attention paid to the criteria for establishing a mechanistic relation. Originally, Bradford-Hill [8], and more recently Kazdin offers [4] at least seven criteria for establishing mechanisms of psychosocial treatments that are equally relevant to implementation strategies:

1 A mediator can point toward a mechanism as it is an intervening variable that may account (statistically) for the relation between the independent variable (strategy) and the dependent variable (implementation outcome), revealing one possible causal pathway for the observed effect [4]. Compared to mediators, mechanisms are conceptualized as more precise in their description of the operations underlying causal processes [5].

| Term       | Definition                                                                 |
|------------|-----------------------------------------------------------------------------|
| Mechanism  | Process or event through which an implementation strategy operates to affect desired implementation outcomes. |
| Precondition| Factor that is necessary in order for an implementation mechanism to be activated. |
| Strategy   | Methods used to promote the implementation of an evidence-based practice or program |
| Determinant| Also commonly referred to as “barriers” and “facilitators,” a factor that enables or hinders the implementation strategy from eliciting the desired effect. |
| Mediator   | Intervening variable that may account for the relationship between the implementation strategy and the implementation outcome. |
| Moderator  | Factor that increase or decrease the level of influence of an implementation strategy. |
| Proximal   | The product of the implementation strategy that is realized because of its specific mechanism of action, the most immediate, observable outcome in the causal pathway. |
| Distal     | Outcome that the implementation processes is ultimately intended to achieve, not the most immediate outcome in the causal pathway. |

| Term       | Definition                                                                 |
|------------|-----------------------------------------------------------------------------|
| Strong association | Association between implementation strategy and mechanism AND between mechanism and behavior change. |
| Specificity | One plausible construct accounts for behavior change. |
| Consistency | Replication of observed results across studies, samples, and conditions. |
| Experimental manipulation | Direct manipulation of implementation strategy or proposed mediator or mechanism shows impact on outcomes. |
| Timeline   | Causes and mediators temporally precede effects and outcomes. |
| Gradient   | Dose response relationship between mediator and outcome. |
| Plausibility or coherence | Explanation invokes other info and steps in a process-outcome relation that are reasonable or supported by other research. |
produce desired effects [4, 8], is important for both research and practice. For research, it is important to specify and examine mechanisms of implementation strategies, especially in the case of null studies, in order to understand why a strategy did or did not achieve its intended effect. For practice, it is crucial to understand mechanisms so that strategies are designed and selected to directly target implementation determinants or barriers. In the absence of this kind of intentional, a priori matching (i.e., strategy targets determinant), it is possible that the “wrong” (or perhaps less potent) strategy will be deployed. This phenomenon of mismatched strategies and determinants was quite prevalent among the 22 tailored improvement intervention studies included in Bosch et al.’s [9] multiple case study analysis. Upon examining the timing of determinant identification and the degree to which included studies informed tailoring of the type versus the content of the strategies using determinant information, they discovered frequent determinant-strategy mismatch across levels of analysis (e.g., clinician-level strategies were used to address barriers that were at the organizational level) [9]. Perhaps what is missing is a clear articulation of implementation mechanisms to inform determinant-strategy matching. We argue that, ultimately, knowledge of mechanisms would help to create a more rational, efficient bundle of implementation strategies that fit specific contextual challenges.

Via a systematic review, we sought to understand how mechanisms are conceptualized and measured, how they are studied (by characterizing the wide array of models and designs used to evaluate mechanisms) and evaluated (by applying Kazdin’s seven criteria), and how much evidence exists for specific mechanisms. In doing so, we offer a rich characterization of the current state of the evidence. In reflecting on this evidence, we provide recommendations for future research to optimize their contributions to mechanistic implementation science.

**Methods**

**Search protocol**

The databases, PubMed and CINAHL Plus, were chosen because of their extensive collection of over 32 million combined citations of medical, nursing and allied health, and life science journals, as well as inclusiveness of international publications. We searched both databases in August 2018 for empirical studies published between January 1990 and August 2018 testing candidate mechanisms of implementation strategies. This starting date was selected given that the concept of evidence-based practice/evidence-based treatment/evidence-based medicine first gained prominence in the 1990’s with the field of implementation science following in response to a growing consciousness of the research to practice gap [10, 11]. The search terms were based on input from all authors who represent a variety of methodological and content expertise related to implementation science and reviewed by a librarian; see Table 3 for all search terms. The search string consisted of three levels with terms reflecting (1) implementation science, (2) evidence-based practice (EBP), and (3) mechanism. We adopted Kazdin’s [4] definition of mechanisms, which he indicates are the basis of an effect. Due to the diversity of definitions that exist in the literature, the term “mechanism” was supplemented with the terms “mediator” and “moderator” to ensure all relevant studies were collected.

**Study inclusion and exclusion criteria**

Studies were included if they were considered an empirical implementation study (i.e., original data collection) and statistically tested or qualitatively explored mechanisms, mediators, or moderators. We did not include dissemination studies given the likely substantive differences between strategies, mechanisms, and outcomes. Specifically, we align with the distinction made between dissemination and implementation put forth by the National Institutes of Health program announcement for Dissemination and Implementation Research in Health that describes dissemination as involving distribution of evidence to a target audience (i.e., communication of evidence) and implementation as involving use of strategies to integrate evidence into target settings (i.e., use of evidence in practice) [12]. However, the word “dissemination” was included in our search terms because of the tendency of some researchers to use “implementation” and “dissemination” interchangeably. Studies were excluded if they were not an implementation study, used the terms “mediator,” “moderator,” or “mechanism” in a different context (i.e., conflict mediator), did not involve the implementation of an EBP, or were a review, concept paper, or opinion piece rather than original research. All study designs were considered. Only

| Table 3 Search strategy |
|-------------------------|
| Search terms | Explanation |
| Implement* OR disseminate* OR "knowledge translation" | These terms were chosen to target Implementation Science literature. |
| AND | |
| "empirically supported treatment" OR "evidence-based practice" OR "evidence-based treatment" OR innovation OR guideline | These terms were chosen to target the implementation evidence-based practices |
| AND | |
| Mediate* OR moderator OR mechanism* | These terms were chosen to target mechanisms explaining the implementation of evidence-based practices |
| NOT | Biological OR microbiology | These terms were chosen to exclude mechanistic studies in biology and microbiology |
studies in English were assessed. See Additional File 1 for exclusion criteria and definitions. We strategically cast a wide net and limited our exclusions so as to characterize the broad range of empirical studies of implementation mechanisms.

Citations generated from the search of PubMed and CINAHL were loaded into EPPI Reviewer 4, an online software program used for conducting literature reviews [13]. Duplicate citations were identified for removal via the duplicate checking function in EPPI and via manual searching. Two independent reviewers (MRB, CWB) screened the first ten citations on title and abstract for inclusion. They then met to clarify inclusion and exclusion criteria with the authorship team, as well as add additional criteria if necessary, and clarify nuances of the inclusion/exclusion coding system (see Additional File 1 for exclusion criteria and definitions). The reviewers met once a week to compare codes and resolve discrepancies through discussion. If discrepancies could not be easily resolved through discussion among the two reviewers, the first author (CCL) made a final determination. During full text review, additional exclusion coding was applied for criteria that could not be discerned from the abstract; articles were excluded at this phase if they only mentioned the study of mechanisms in the discussion or future directions. Seven studies from the previous systematic review of implementation mechanisms [14] were added to our study for data extraction; these studies likely did not appear in our review due to differences in the search strategy in that the review undertaken by Williams hand searched published reviews of implementation strategies in mental health.

Study quality assessment
The methodological quality of included studies was assessed using the Mixed Methods Appraisal Tool (MMAT-version 2018) [15]. This tool has been utilized in over three dozen systematic reviews in the health sciences. The MMAT includes two initial screening criteria that assess for the articulation of a clear research question/objective and for the appropriateness of the data collected to address the research question. Studies must receive a “yes” in order to be included. The tool contains a subset of questions to assess for quality for each study type—qualitative, quantitative, and mixed methods. Table 4 summarizes the questions by which studies were evaluated, such as participant recruitment and relevance and quality of measures. Per the established approach to MMAT application, a series of four questions specific to each study design type are assigned a dichotomous “yes” or “no” answer. Studies receive 25 percentage points for each “yes” response. Higher percentages reflect higher quality, with 100% indicating all quality criteria were met. The MMAT was applied by the third author (CWB). The first author (CCL) checked the first 15% of included studies and, based on reaching 100% agreement on the application of the rating criteria, the primary reviewer then applied the tool independently to the remaining studies.

Data extraction and synthesis
Data extraction focused on several categories: study information/ background (i.e., country, setting, and sample), methods (i.e., theories that informed study, measures used, study design, analyses used, proposed mediation model), results (i.e., statistical relations between proposed variables of the mediation model tested), and criteria for establishing mechanisms (based on the seven listed in Table 2 [4]). All authors contributed to the development of data extraction categories that were applied to the full text of included studies. One reviewer (MRB) independently extracted relevant data and the other reviewer (CWB) checked the results for accuracy, with the first author (CCL) addressing any discrepancies or questions, consistent with the approach of other systematic reviews [61]. Extracted text demonstrating evidence of study meeting (or not meeting) each criterion for establishing a mechanism was further independently coded as “1” reflecting “criterion met” or “0” reflecting “criterion not met” by MRB and checked by CWB. Again, discrepancies and questions were resolved by the first author (CCL). Technically, mechanisms were considered “established” if all criteria were met. See Additional File 2 for PRISMA checklist for this study.

Results
The search of PubMed and CINAHL Plus yielded 2277 studies for title and abstract screening, of which 447 were duplicates, and 183 moved on to full-text review for eligibility. Excluded studies were most frequently eliminated due to the use of mechanism in a different context (i.e., refer to a process, technique, or system for achieving results of something other than implementation strategies). After full article review, 39 studies were deemed suitable for inclusion in this review. Two of the included studies appeared in the only other systematic review of implementation mechanisms in mental health settings [14]. For consistency and comprehensiveness, the remaining seven studies from the previously published review were added to the current systematic review for a total of 46 studies. See Fig. 1 for a PRISMA Flowchart of the screening process and results.

2Key differences in Williams’ [14] search method are important to note. Williams first conducted a broad search for randomized controlled trials concerning implementation or dissemination of evidence-based therapies. Only after screening references for these criteria, did Williams narrow the search to studies that specifically addressed mediators. Conversely, the present method included mediators/moderators/mechanisms as terms in the initial search string. Additionally, Williams hand searched references included in four previous reviews of implementation strategies in mental health.
Table 4 MMAT

| Data source relevant? | Data analysis process relevant? | Findings relate to context? | Findings relate to researchers’ influence? | Total score (%) |
|-----------------------|----------------------------------|----------------------------|-------------------------------------------|----------------|
| Y                     | Y                                | Y                          | N                                         | 75             |
| Y                     | Y                                | Y                          | Y                                         | 75             |
| Y                     | Y                                | Y                          | Y                                         | 75             |
| Y                     | Y                                | Y                          | Y                                         | 75             |
| N/A                   | N/A                              | N/A                        | N/A                                       | 50             |
| N/A                   | N/A                              | N/A                        | N/A                                       | 50             |
| N/A                   | N/A                              | N/A                        | N/A                                       | 50             |
| N/A                   | N/A                              | N/A                        | N/A                                       | 50             |

3. Quantitative - non-randomized

| Recruitment minimizes selection bias? | Measurements appropriate? | Comparable groups or control for differences? |
|---------------------------------------|---------------------------|-----------------------------------------------|
| Y                                     | Y                         | Y                                             |
| Y                                     | Y                         | Y                                             |
| Y                                     | Y                         | Y                                             |
| Y                                     | Y                         | Y                                             |
| Y                                     | Y                         | Y                                             |

4. Qualitative

| Data source relevant? | Data analysis process relevant? | Findings relate to context? | Findings relate to researchers’ influence? | Total score (%) |
|-----------------------|----------------------------------|----------------------------|-------------------------------------------|----------------|
| Y                     | Y                                | Y                          | N                                         | 75             |
| Y                     | Y                                | Y                          | Y                                         | 75             |
| Y                     | Y                                | Y                          | Y                                         | 75             |
| Y                     | Y                                | Y                          | Y                                         | 75             |
| N/A                   | N/A                              | N/A                        | N/A                                       | 50             |
| N/A                   | N/A                              | N/A                        | N/A                                       | 50             |
| N/A                   | N/A                              | N/A                        | N/A                                       | 50             |
| N/A                   | N/A                              | N/A                        | N/A                                       | 50             |

3. Quantitative - non-randomized

| Recruitment minimizes selection bias? | Measurements appropriate? | Comparable groups or control for differences? |
|---------------------------------------|---------------------------|-----------------------------------------------|
| Y                                     | Y                         | Y                                             |
| Y                                     | Y                         | Y                                             |
| Y                                     | Y                         | Y                                             |
| Y                                     | Y                         | Y                                             |
| Y                                     | Y                         | Y                                             |
| Y                                     | Y                         | Y                                             |
| Y                                     | Y                         | Y                                             |

4. Qualitative

| Data source relevant? | Data analysis process relevant? | Findings relate to context? | Findings relate to researchers’ influence? | Total score (%) |
|-----------------------|----------------------------------|----------------------------|-------------------------------------------|----------------|
| Y                     | Y                                | Y                          | N                                         | 75             |
| Y                     | Y                                | Y                          | Y                                         | 75             |
| Y                     | Y                                | Y                          | Y                                         | 75             |
| Y                     | Y                                | Y                          | Y                                         | 75             |
| N/A                   | N/A                              | N/A                        | N/A                                       | 50             |
| N/A                   | N/A                              | N/A                        | N/A                                       | 50             |
| N/A                   | N/A                              | N/A                        | N/A                                       | 50             |
| N/A                   | N/A                              | N/A                        | N/A                                       | 50             |

3. Quantitative - non-randomized

| Recruitment minimizes selection bias? | Measurements appropriate? | Comparable groups or control for differences? |
|---------------------------------------|---------------------------|-----------------------------------------------|
| Y                                     | Y                         | Y                                             |
| Y                                     | Y                         | Y                                             |
| Y                                     | Y                         | Y                                             |
| Y                                     | Y                         | Y                                             |
| Y                                     | Y                         | Y                                             |
| Y                                     | Y                         | Y                                             |
| Y                                     | Y                         | Y                                             |
| Y                                     | Y                         | Y                                             |

4. Qualitative

| Data source relevant? | Data analysis process relevant? | Findings relate to context? | Findings relate to researchers’ influence? | Total score (%) |
|-----------------------|----------------------------------|----------------------------|-------------------------------------------|----------------|
| Y                     | Y                                | Y                          | N                                         | 75             |
| Y                     | Y                                | Y                          | Y                                         | 75             |
| Y                     | Y                                | Y                          | Y                                         | 75             |
| Y                     | Y                                | Y                          | Y                                         | 75             |
| N/A                   | N/A                              | N/A                        | N/A                                       | 50             |
| N/A                   | N/A                              | N/A                        | N/A                                       | 50             |
| N/A                   | N/A                              | N/A                        | N/A                                       | 50             |
| N/A                   | N/A                              | N/A                        | N/A                                       | 50             |

3. Quantitative - non-randomized

| Recruitment minimizes selection bias? | Measurements appropriate? | Comparable groups or control for differences? |
|---------------------------------------|---------------------------|-----------------------------------------------|
| Y                                     | Y                         | Y                                             |
| Y                                     | Y                         | Y                                             |
| Y                                     | Y                         | Y                                             |
| Y                                     | Y                         | Y                                             |
| Y                                     | Y                         | Y                                             |
| Y                                     | Y                         | Y                                             |
| Y                                     | Y                         | Y                                             |
| Y                                     | Y                         | Y                                             |
| Table 4 MMAT (Continued) |
|--------------------------|
| Complete outcome data, acceptable response rate, or acceptable follow-up rate? | N | N | N | N | N | N | N | N | Y | N | N | N | N | N | N | N | N | N | N | N |
| Total score (%) | 75 | 50 | 50 | 50 | 50 | 75 | 25 | 75 | 75 | 75 | 75 | 100 | 50 | 75 | 75 | 75 | 75 |
| Armson et al. 2018 [54] | Y | N | N | N | N | N | Y | N | N | N | N | N | N | N | N | N |
| Birken et al. 2015 [55] | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Kauth et al. 2010 [56] | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Lukas et al. 2009 [57] | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Panzano et al. 2012 [58] | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Rangachari et al. 2015 [59] | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Shrubsole et al. 2018 [60] | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |

1. Qualitative
- Data sources relevant?
- Data analysis process relevant?
- Findings relate to context?
- Findings relate to researchers' influence?

2. Quantitative randomized
- Clear description of the randomization?
- Clear description of allocation or concealment?
- Complete outcome data?
- Low withdrawal/drop-out?

3. Quantitative non-randomized
- Recruitment minimizes selection bias?
- Measurements appropriate?
- Comparable groups or control for differences?
- Complete outcome data, acceptable response rate, or acceptable follow-up rate?

4. Mixed methods
- Research design relevant?
Table 4 MMAT (Continued)

| Integration of qualitative and quantitative data relevant? | Y | Y | Y | Y | Y | Y | Y |
|------------------------------------------------------------|---|---|---|---|---|---|---|
| Appropriate consideration given to limitations associated with integration? | Y | Y | N | N | N | N | N |
| Total score (%) | 75 | 50 | 50 | 25 | 25 | 75 | 75 |
Study characteristics

**Setting, sampling, and interventions**

Table 5 illustrates the characteristics of the 46 included studies. Twenty-five studies (54.3%) were completed in the USA, while 21 studies were conducted in other countries (e.g., Australia, Canada, Netherlands, UK). Settings were widely variable; studies occurred in behavioral health (e.g., community mental health, residential facilities) or substance abuse facilities most frequently (21.7%), followed by hospitals (15.2%), multiple sites across a health care system (15.2%), schools (15.2%), primary care clinics (10.9%), and Veteran’s Affairs facilities (8.7%). Sampling occurred at multiple ecological levels, including patients (17.4%), providers (65.2%), and organizations (43.5%). Seventeen (40.0%) studies examined the implementation of a complex psychosocial intervention (e.g., Cognitive behavioral therapy [42, 56]; multisystemic therapy [25, 26, 58]).

**Study design**

Our review included six qualitative (10.9%), seven mixed methods (15.2%), and 34 quantitative studies (73.9%). The most common study design was quantitative non-randomized/observational (21 studies; 45.7%), of which 11 were cross-sectional. There were 13 (28.3%) randomized studies included in this review. Twenty-nine studies
| Study | Setting | Sample | Intervention/Innovation | Complex psychosocial intervention | Design |
|-------|---------|--------|-------------------------|----------------------------------|--------|
| Bardosh et al. 2017 [16] | Health care facilities, multiple countries | Key informants (researchers, Mhealth staff, clinic staff, government officials; n = 32) | Mobile health application | N | Qualitative, cross sectional, comparative case study, non-randomized |
| Brewster et al. 2015 [17] | Hospitals | Hospitals (k = 10); hospital employees (hospital staff, n = 82; state hospital representatives n = 8) | Initiative to reduce rehospitalization rates | N | Qualitative, descriptive, cross sectional, non-randomized |
| Carrera and Lambooij 2015 [18] | Primary care | Patients (n = 12); health care providers (n = 4) | Blood pressure monitoring guidelines | N | Qualitative descriptive, cross sectional, non-randomized |
| Frykman et al. 2014 [19] | Emergency departments | Departments (k = 2), health care providers (n = 11) | Multi-professional teamwork guideline | N | qualitative, longitudinal (2 assessment points, 21 months), comparative case study, non-randomized |
| Wiener-Ogilvie et al. 2008 [20] | Primary care | Health care providers (n = 9) | Asthma management guideline | N | qualitative, cross sectional, comparative case study, non-randomized |
| Atkins et al. 2008 [21] | Schools | Teachers (n = 127); mental health providers (n = 21) | Attention Deficit Hyperactivity Disorder guidelines | Y | quantitative, longitudinal (5 assessment points, 2 years), randomized |
| Baer et al. 2009 [22] | Substance abuse treatment facilities | Substance abuse treatment facilities (k = 6); Mental health providers (n = 118) | Motivational Interviewing | Y | quantitative, longitudinal (3 assessment points, 6 months), randomized |
| Bonetti et al. 2005 [23] | Primary care | Health care providers (n = 152) | Spinal X-ray referral guidelines | N | quantitative, longitudinal (2 assessment points, 2 months), randomized control trial |
| Garner et al. 2011 [24] | Substance abuse treatment facilities | Substance abuse treatment facilities (k = 29); mental health providers (n = 95) | Adolescent Community Reinforcement Approach and Assertive Continuing Care | Y | quantitative, longitudinal (2 assessment points, 3 years), randomized control trial |
| Glisson et al. 2010 [25] | Juvenile courts | Counties (k = 14); patients (n = 615) | Multisystemic Therapy | Y | quantitative, longitudinal (weekly, quarterly, 4 years), randomized control trial |
| Holth et al. 2011 [26] | Behavioral health facilities | Mental health providers (n = 21); families (youth and primary caregiver; n = 41) | Multisystemic Therapy, Cognitive Behavior Therapy | Y | quantitative, longitudinal (monthly, 17 months), block randomized control trial |
| Lee et al. 2018 [27] | Schools, child care facilities | Organizations (n = 121) | Nutritional guidelines | N | quantitative, longitudinal (two time points; 2 studies at 6 months, 1 study at 12 months), analysis of aggregated datasets from three randomized control trials |
| Lochman et al. 2009 [28] | Schools | Schools (k = 57); patients (n = 531); mental health providers (n = 49) | Coping Power Program | Y | quantitative, longitudinal (2 assessment points, 2 years), randomized |
| Rapkin et al. 2017 [29] | Public library system | Communities (k = 20); community members (n = 9374) | Cancer screening and prevention education programs | N | quantitative, randomized, stepped-wedge, longitudinal |
| Rohrbach et al. 1993 [30] | Schools | Schools (k = 25); administrators (n = 25); teachers (n = 60); patients (n = 1147) | Adolescent Alcohol Prevention Trial | Y | quantitative, longitudinal (3 assessment points, 2 years), randomized control trial |
### Table 5: Descriptive summary (Continued)

| Study                                      | Setting                  | Sample                                                                 | Intervention/Innovation                                      | Complex psychosocial intervention | Design                                                                 |
|--------------------------------------------|--------------------------|------------------------------------------------------------------------|-------------------------------------------------------------|-----------------------------------|------------------------------------------------------------------------|
| Seys et al. 2018 [31]                      | Hospitals                | Care teams (k = 19); care team members (n = 284); patients (n = 257) | Care pathway for Chronic Obstructive Pulmonary Disease      | N                                 | quantitative, longitudinal (two assessment points, 30 days), randomized |
| Williams et al. 2014 [32]                  | Behavioral health facilities | Behavioral health facilities (k = 92); administrators (n = 311)       | Motivational Interviewing                                   | Y                                 | quantitative, longitudinal (3 assessment points, 3 months), randomized control trial |
| Williams et al. 2017 [33]                  | Behavioral health facilities | Organizations (k = 14); clinicians (n = 475)                  | Evidence-based practice (not specified)                    | Evidence-based practice implemented not reported | quantitative, longitudinal, randomized (4 assessment points, 4 years) |
| Quantitative non-randomized                |                          |                                                                        |                                                             |                                   |                                                                        |
| Aarons et al. 2009 [34]                    | Behavioral health facilities | Mental health care providers (n = 174)                                 | 31 child or family evidence-based practices                 | Y*                                | quantitative, cross-sectional, survey, non-randomized                  |
| Becker et al. 2016 [35]                    | Substance abuse treatment facilities | Clinics (k = 15); treatment providers (n = 60)                   | Contingency management treatment                           | Y                                 | quantitative, longitudinal (biweekly, 12 months), non-randomized        |
| Beenstock et al. 2012 [36]                 | Hospitals                | Hospitals (k = 8); health care providers (n = 364)                  | Smoking cessation guideline                                 | N                                 | quantitative, cross sectional, survey, non-randomized                  |
| Beets et al. 2008 [37]                     | Schools                  | Teachers (n time 1 = 171, n time 2 = 191)                            | Positive Action Program                                     | Y                                 | quantitative, cross sectional at two time points, survey, non-randomized |
| Bonetti et al. 2009 [38]                   | Dentist offices          | Health care providers (n = 133)                                      | Fissure sealant evidence-based practice                    | N                                 | quantitative, longitudinal, predictive cohort study (3 assessment points, 28 months), non-randomized |
| Chou et al. 2011 [39]                      | Veterans Affairs         | Hospitals (k = 132), health care providers (n = 2,438)              | Major depressive disorder screening guideline               | N                                 | quantitative, cross sectional, survey, randomized                      |
| Cummings et al. 2017 [40]                 | Nursing homes            | Nursing homes (k = 7); nursing home staff (n = 333)                | Coaching for Impressive Care                               | N                                 | quantitative, , non-randomized two-group crossover                     |
| David and Schiff 2017 [41]                 | Health care system, multiple sites | Health care providers (n = 77)                                      | Child-Parent Psychotherapy                                  | Y                                 | quantitative, cross sectional, survey, non-randomized                  |
| Edmunds et al. 2014 [42]                   | Behavioral health facilities | Mental health providers (n = 50)                                      | Cognitive Behavioral Therapy                                | Y                                 | quantitative, longitudinal, non-randomized                              |
| Gnich et al. 2018 [43]                     | Dentist offices          | Health care providers (n = 709)                                      | Fluoride varnish application                               | N                                 | quantitative, longitudinal (2 assessment points, 18 months), non-randomized |
| Guerrero et al. 2018 [44]                  | Behavioral health facilities | Behavioral health facilities (k = 112), mental heal providers (n = 427) | Contingency management treatment and medication assisted treatment | Y                                 | quantitative, longitudinal (2 assessment points), survey, non-randomized |
| Huis et al. 2013 [45]                      | Hospitals                | Hospitals (k = 3); departments (k = 677); health care providers (k = 2733) | Hand hygiene guidelines                                    | N                                 | quantitative, longitudinal, process evaluation of a cluster randomized controlled trial |
| Little et al. 2015 [46]                    | Schools                  | School districts (k = 183); departments (k = 22)                   | Tobacco Use Prevention Education                            | N                                 | quantitative, cross sectional, survey, non-randomized                  |
| Llasus et al. 2014 [47]                    | University nursing programs | Nursing students (n = 174)                                         | Evidence-based practices (not specified)                    | N                                 | quantitative, descriptive, correlational, cross sectional, survey, non-randomized |
| Study | Setting | Sample | Intervention/Innovation | Complex psychosocial intervention | Design |
|-------|---------|--------|-------------------------|----------------------------------|-------|
| Nelson and Steele 2007 [48] | Health care system, multiple sites | Mental health providers (n = 214) | Evidence-based practices (not specified) | N | quantitative, cross sectional, survey, non-randomized |
| Potthoff et al. 2017 [49] | Primary care | Organizations (k = 99); health care providers (n = 489) | Type 2 diabetes management guideline | N | quantitative, longitudinal (2 assessment points, 1 year), correlational, survey, non-randomized |
| Presseau et al. 2016 [50] | Primary care | Family physicians (time 1 n = 632; time 2 n = 426) | Prescription of hypertension medication | N | quantitative, longitudinal (2 assessment points, approximately 8 months), 2X3 factorial |
| Simmonds et al. 2012 [51] | Health care system, multiple sites | Health care providers (n = 108) | Lower back pain management guidelines | N | quantitative, cross sectional, survey, non-randomized |
| Stockdale et al. 2018 [52] | Veterans Affairs | Health care providers (n = 149), patients (n = 3329) | Patient Centered Medical Home | N | quantitative, cross sectional, survey, non-randomized |
| Wanless et al. 2015 [53] | Schools | Schools (k = 13); teachers (n = 1114) | Responsive Classroom | Y | quantitative, longitudinal, non-randomized (focuses on one condition in an RCT) |
| Yamada et al. 2017 [62] | Hospitals | Care units (k = 32); nurses (n = 779); patients (n = 1,604) | Instrumental and conceptual research use, evidence-based pain assessment | N | quantitative, cross sectional, non-randomized |

**Mixed Methods**

| Study | Setting | Sample | Intervention/Innovation | Complex psychosocial intervention | Design |
|-------|---------|--------|-------------------------|----------------------------------|-------|
| Armson et al. 2018 [54] | Health care system, multiple sites | Health care providers (n = 70) | Breast cancer screening guideline | N | mixed method, longitudinal, observational/ naturalist field study, non-randomized |
| Birken et al. 2015 [55] | Health care system, multiple sites | Organizations (k = 149); administrators (n = 223) | Quality improvement initiative based on Chronic Care Model | N | mixed method sequential, cross sectional, non-randomized |
| Kauth et al. 2010 [56] | Veterans Affairs | Clinics (k = 21); mental health providers (n = 23) | Cognitive Behavioral Therapy | Y | mixed method, quasi-experimental, longitudinal (2 assessment points, 6 months), randomized |
| Lukas et al. 2009 [57] | Veterans Affairs | Organizations (k = 78); health care providers, non-clinical staff (n = 3870) | Advance Clinic Access | N | mixed method, cross sectional, observational, non-randomized |
| Panzano et al. 2012 [58] | Behavioral health facilities | Consultants (n = 34); mental health providers (n = 70) | Multisystemic Therapy, Dual Disorder Treatment, Ohio medication algorithms, Cluster-based Outcomes Management | Y | mixed method, longitudinal, observational/ naturalist field study, non-randomized |
| Rangachari et al. 2015 [59] | Hospitals | Departments (k = 2); health care providers (n = 101); administrators (n = 6) | Central line bundle catheter insertion evidence-based practice | N | prospective, longitudinal, exploratory field study, mixed-method analysis |
| Shrubsole et al. 2018 [60] | Hospitals | Hospitals (k = 4); health care providers (n = 37); patients (n = 107) | Aphasia management practices | N | mixed method, longitudinal, cross-over, cluster randomized control trial |

*Multiple EBPs, some of which were complex psychosocial interventions*
(63.0%) were longitudinal (i.e., included more than one data collection time point for the sample).

**Study quality**
Table 4 shows the results of the MMAT quality assessment. Scores for the included studies ranged from 25 to 100%. Six studies (13.0%) received a 25% rating based on the MMAT criteria [15], 17 studies (40.0%) received 50%, 21 studies (45.7%) received 75%, and only three studies (6.5%) scored 100%. The most frequent weaknesses were the lack of discussion on researcher influence in qualitative and mixed methods studies, lack of clear description of randomization approach utilized in the randomized quantitative studies, and subthreshold rates for acceptable response or follow-up in non-randomized quantitative studies.

**Study design and evaluation of mechanisms theories, models, and frameworks**
Twenty-seven (58.7%) of the studies articulated their plan to evaluate mechanisms, mediators, or moderators in their research aims or hypotheses; the remaining studies included this as a secondary analysis. Thirty-five studies (76.1%) cited a theory, framework, or model as the basis or rationale for their evaluation. The diffusion of innovations theory [63, 64] was most frequently cited, appearing in nine studies (19.6%), followed by the theory of planned behavior [65], appearing in seven studies (15.2%). The most commonly cited frameworks were the theoretical domains framework (five studies; 10.9%) [66] and Promoting Action on Research in Health Services (PARiHS) [67] (three studies; 6.5%).

**Ecological levels**
Four studies (8.7%) incorporated theories or frameworks that focused exclusively on a single ecological level; two focusing on leadership, one on the organizational level, and one at the systems level. There was some discordance between the theories that purportedly informed studies and the potential mechanisms of interest, as 67.4% of candidate mechanisms or mediators were at the intrapersonal level, while 30.4% were at the interpersonal level, and 21.7% at the organizational level. There were no proposed mechanisms at the systems or policy level. Although 12 studies (26.1%) examined mechanisms or mediators across multiple ecological levels, few explicitly examined multilevel relationships (e.g., multiple single-level mediation models were tested in one study).

**Measurement and analysis**
The vast majority of studies (38, 82.6%) utilized self-report measures as the primary means of assessing the mechanism, and 13 of these studies (28.3%) utilized focus groups and/or interviews as a primary measure, often in combination with other self-report measures such as surveys. Multiple regression constituted the most common analytic approach for assessing mediators or moderators, utilized by 25 studies (54.3%), albeit this was applied in a variety of ways. Twelve studies (26.1%) utilized hierarchical linear modeling (HLM) and six studies (13.0%) utilized structural equation modeling (SEM); see Table 6 for a complete breakdown. Studies that explicitly tested mediators employed diverse approaches including Baron and Kenny’s (N = 8, 17.4 causal steps approach [78], Preacher and Hayes’ (N = 3, 6.5%) approach to conducting bias-corrected bootstrapping to estimate the significance of a mediated effect (i.e., computing significance for the product of coefficients) [95, 126], and Sobel’s (N = 4, 8.9%) approach to estimating standard error for the product of coefficients often using structural equation modeling [79]. Only one study tested a potential moderator, citing Raudenbush’s [80, 82]. Two other studies included a potential moderator in their conceptual frameworks, but did not explicitly test moderation.

**Emergent mechanism models**
There was substantial variation in the models that emerged from the studies included in this review. Table 7 represents variables considered in mediating or moderating models across studies (or identified as candidate mediators, moderators, or mechanisms in the case of qualitative studies). Additional file 3 depicts the unique versions of models tested and their associated studies. We attempted to categorize variables as either (a) an independent variable (X) impacting a dependent variable; (b) a dependent variable (Y), typically the outcome of interest for a study; or (c) an intervening variable (M), a putative mediator in most cases, though three studies tested potential moderators. We further specified variables as representing a strategy, determinant, and outcome; see Table 1 for definitions.3

**Common model types**
The most common model type (29; 63.0%) was one in which X was a determinant, M was also a determinant, and Y was an implementation outcome variable (determinant ➔ determinant ➔ implementation outcome). For example, Beenstock et al. [36] tested a model in which propensity to act (determinant) was evaluated as a mediator explaining the relation between main place of work (determinant) and referral to smoking cessation services (outcome). Just less than half the studies (22; 47.8%)

3We refer to variables in the ways the study authors did, even if we might have a different way in which we would approach their conceptualization.
| Study                      | Aims           | Theory, framework, model                                                                 | Mechanism measurement                        | Mediation testing citation |
|---------------------------|----------------|------------------------------------------------------------------------------------------|-----------------------------------------------|-----------------------------|
| **Qualitative**          |                |                                                                                          |                                               |                             |
| Bardosh et al. 2017 [16]  | N              | Consolidated framework for implementation research [58]                                  | Interviews                                   | None                        |
| Brewster et al. 2015 [17] | Y              | Implementation innovation framework [69]                                                  | Interviews                                   | None                        |
| Carrera and Lambooij 2015 | N              | Technology acceptance model [70]; Theory of planned behavior [65]; Model of personal computing utilization [71] | Focus groups                                 | None                        |
| Frykman et al. 2014 [18]  | N              | Direction, competence, opportunity and motivation (DCOM) [72, 73]                        | Interviews; observations                     | None                        |
| Wiener-Ogilvie et al. 2008 [20] | N        | None reported                                                                            | Interviews; focus groups                     | None                        |
| **Quantitative- randomized** |                |                                                                                          |                                               |                             |
| Atkins et al. 2008 [21]   | Y              | Diffusion of innovation theory [63]                                                        | Interviews; self-report                       | [74]                        |
| Baer et al. 2009 [22]      | Y              | None reported                                                                            | interviews; self-report                      | [75]                        |
| Bonetti et al. 2005 [23]   | N              | Theory of planned behavior [65]; Social cognitive theory [76, 77]                         | Self-report                                  | [78, 79]                    |
| Garner et al. 2011 [24]    | N              | Theory of planned behavior [65]                                                            | Self-report                                  | [80, 81]                    |
| Glisson et al. 2010 [25]   | N              | None reported                                                                            | Self-report, audiotape coding and interviews  | [82]                        |
| Holth et al. 2011 [26]     | Y              | None reported                                                                            | Interviews; self-report                      | [83]                        |
| Lee et al. 2018 [27]       | Y              | Theoretical domains framework [84]                                                        | Self-report, secondary analysis              | [85, 86]                    |
| Lochman et al. 2009 [28]    | N              | Diffusion of innovation theory [87]                                                        | Coder ratings                                | [88]                        |
| Rapkin et al. 2017 [29]    | Y              | None reported                                                                            | Self-report                                  | [89]                        |
| Rohrbach et al. 1993 [30]  | N              | Diffusion of innovation theory [64]                                                        | Interviews; self-report; observations        | None                        |
| Seys et al. 2018 [31]      | Y              | None reported                                                                            | Chart review; self-report                    | [78]                        |
| Williams et al. 2014 [32]  | Y              | Diffusion of innovation theory [87]                                                        | Self-report                                  | [90, 91]                    |
| Williams et al., 2017 [66] | Y              | Organizational culture theory [32] and Theory of planned behavior [65]                    | Self-report                                  | [92]                        |
| **Quantitative- non-randomized** |              |                                                                                          |                                               |                             |
| Aarons et al. 2009 [34]    | Y              | Institutional theory [93], Theory of planned behavior [65], Theory of perceived organizational support [94] | Self-report                                  | [78]                        |
| Becker et al. 2016 [35]    | Y              | Diffusion of innovation theory [64]                                                        | Self-report                                  | None                        |
| Beenstock et al. 2012 [36] | N              | Theoretical domains framework [66]                                                        | Self-report                                  | [95]                        |
| Beets et al. 2008 [37]      | Y              | Theory driven evaluation [96]; Diffusion of innovation theory [63]                        | Self-report                                  | [97, 98]                    |
| Bonetti et al. 2009 [38]    | N              | Theory of planned behavior [65]; Social cognitive theory [99]; Operant learning theory [100]; Action planning [101]; Common sense self-regulation model [102]; Precaution adoption process model [103]; Stage theory [103, 104] | Self-report; objective measure              | [78, 79]                    |
| Chou et al. 2011 [39]       | N              | Goal setting theory [105]; Goal commitment theory [106]                                   | Self-report                                  | [80, 107]                   |
included an implementation strategy in their model, of which 16 (34.8%) evaluated a mediation model in which an implementation strategy was $X$, a determinant was the candidate $M$, and an implementation outcome was $Y$ (strategy $\rightarrow$ determinant $\rightarrow$ implementation outcome); ten of these studies experimentally manipulated the relation between the implementation strategy and determinant. An example of this more traditional mediation model is a study by Atkins and colleagues [21] which evaluated key opinion leader support and mental health practitioner support (determinants) as potential mediators of the relation between training and consultation.

Table 6 Mechanism analysis (Continued)

| Study                        | Aims | Theory, framework, model                              | Mechanism measurement           | Mediation testing citation |
|------------------------------|------|-------------------------------------------------------|---------------------------------|----------------------------|
| Cummings et al. 2017 [40]    | N    | Promoting action on research in health services (PARiHS) [67] | Self-report                      | [108]                      |
| David and Schiff 2017 [41]   | Y    | Diffusion of innovation theory [87, 109]              | Self-report                      | [110]                      |
| Edmunds et al. 2014 [42]     | Y    | EPIS framework [111]                                 | Self-report                      | [80, 112]                  |
| Gnich et al. 2018 [43]       | Y    | Theoretical domains framework [66]                   | Self-report                      | None                       |
| Guerrero et al. 2018 [44]    | Y    | Theory on middle manager s[69]                       | Self-report                      | [113]                      |
| Huis et al. 2013 [45]        | N    | None reported                                         | Observations; self-report; website visitor registration; logs; field Notes; effect evaluation; quiz | none |
| Little et al. 2015 [46]      | N    | Diffusion of innovation theory [64]                  | Self-report                      | [114–116]                  |
| Llussa et al. 2014 [47]      | N    | Knowledge to action conceptual framework [117]       | Self-report                      | [78, 79, 95]               |
| Nelson and Steele 2007 [48]  | N    | None reported                                         | Self-report                      | None                       |
| Poitthoff et al. 2017 [49]   | Y    | Dual process model of behavior [118]                 | Self-report                      | [79]                       |
| Presseau et al. 2016 [50]    | Y    | Theory of planned behavior [65]                      | Self-report                      | None                       |
| Simmons et al. 2012 [51]     | Y    | None reported                                         | Self-report                      | [78]                       |
| Stockdale et al. 2018 [52]   | Y    | None reported                                         | Self-report                      | [119]                      |
| Wanless et al. 2015 [53]     | Y    | None reported                                         | Self-report, observation         | [110]                      |
| Yamada et al. 2017 [54]      | Y    | Promoting action on research in health services (PARiHS) [120] | Self-report, chart review       | None                       |
| Mixed methods                |      |                                                       |                                 |                            |
| Armson et al. 2018 [54]      | Y    | Theoretical domains framework [66]                   | Interviews; self-report          | None                       |
| Birken et al. 2015 [55]      | N    | Hierarchical taxonomy of leader behavior [121]       | Interviews; self-report          | [95, 122]                  |
| Kauth et al. 2010 [56]       | Y    | Fixsen model [123]; Promoting action on research in health services (PARiHS) [120] | Self-report; logs               | None                       |
| Lukas et al. 2009 [57]       | Y    | Diffusion of Innovations Theory [63, 124]            | Interviews                       | [78]                       |
| Panzano et al. 2012 [58]     | Y    | None reported                                         | Self-report                      | [78]                       |
| Rangachari et al. 2015 [59]  | N    | Complex systems theory [125]                          | Infection rate; chart review; hospital records; logs | None                       |
| Shrubsole et al. 2018 [60]   | N    | Theoretical domains framework [66]                   | Chart review; self-report        | none                       |
| Study | Independent variable (X) | Intervening variable (M) | Dependent variable (Y) |
|-------|--------------------------|--------------------------|------------------------|
| **Qualitative** | | | |
| Bardosh et al. 2017 [16] | Mobile and text follow up with patients | Service organization at clinic level, clinician norms and practices, availability of local champions staff, adaptability and co-design of strategy, receptivity and capacity of local management | Culture of care |
| Brewster et al. 2015 [17] | Patient education, follow-up phone calls to patients after discharge, discharge planning, collaboration with post-acute providers | Intrinsic reward to staff --> shift in norms and attitudes | Reduced hospital readmissions |
| Carrera and Lambooij 2015 [18] | None reported | Mediators: perceived usefulness, perceived ease of use, self-efficacy, attitudes, social norm | Intervention acceptability (providers and patients) |
| Frykman et al. 2014 [19] | Senior manager and consultant-driven teamwork strategy, senior manager and staff-driven teamwork strategy | Direction, communication, opportunity, motivation | Change in staff behavior |
| Wiener-Ogilvie et al. 2008 [20] | Guideline implementation | Practice organization (delegation of work to nurses) | Compliance with guidelines |
| **Quantitative—randomized** | | | |
| Atkins et al. 2008 [21] | Training and consultation | Key opinion leader instrumental supportmental health professional instrumental support | Teacher self-reported used of ADHD guidelines |
| Baer et al. 2009 [22] | Climate for organizational change | Post training agency activities to support use of Motivational Interviewing | Fidelity to intervention (Motivational Interviewing spirit and response to question ratio) |
| Bonetti et al. 2005 [23] | Audit and feedback | Decision difficulty, behavioral control | Simulated behavior |
| Garner et al. 2011 [24] | Pay for performance | 1. Subjective norms 2. Attitudes toward intervention 3. Perceived control | 1. Therapists’ intention to achieve monthly competence 2. Therapists’ intention to achieve targeted threshold |
| Glisson et al. 2010 [25] | Availability responsiveness and continuity (ARC) Intervention + Multisystemic Therapy quality assurance, pay for performance | Fidelity to multisystemic therapy | Rate of change in child behavior out of home placements |
| Holth et al. 2011 [26] | Workshop + manual, intensive quality assurance + workshop + manual | Adherence to contingency management and cognitive behavioral therapy techniques | Youth cannabis use |
| Lee et al. 2018 [27] | Implementation strategy bundles (varied across studies) | Knowledge, skills, social/professional role and identity, environmental resources | Nutrition guideline implementation |
| Lochman et al. 2009 [28] | Intensive training + feedback, basic training | # of sessions attended, # of objectives completed, # of contacts with trainers, counselor engagement w/clients | Client externalizing behaviors, client social skills, client study skills, client expectancies re: aggression, consistent parenting, client assaultive acts |
| Rapkin et al. 2017 [29] | Indicators of program activities: cumulative local programs, attendance at local programs, time since most recent local program, personal awareness of programs, cumulative outside programs | Mediators: awareness of free/low cost cancer screening, cancer knowledge, cancer information seeking, having health insurance, annual physical moderator: frequency of library use | Cancer screening attempts to quit smoking/tobacco cessation |
| Rohrbach et al. 1993 [30] | 1. Teacher training 2. Principal support intervention | 1a. Teacher self-efficacy, 1b. enthusiasm, 1c. preparedness 2a. Principal encouragement, 2b. Principal beliefs about program | Quantity of program implementation |
| Seys et al. 2018 [31] | Care pathway implementation | Adherence to evidence-based recommendations, level of competence, team climate for innovation, burnout, level of organized care | 30-day hospital readmission |
| Williams et al. 2014 [32] | Information packets and Motivational Interviewing webinar | Attitudes towards EBPs, pressure for change, barriers to EBPs, resources, organizational climate, management support | Motivational Interviewing adoption |
| Williams et al. 2017 [33] | Availability, Responsiveness, and Continuity (ARC) intervention implementation | Proficiency culture --> evidence-based practice intention, barrier reduction | EBP adoption, EBP use |
| Study | Independent variable (X) | Intervening variable (M) | Dependent variable (Y) |
|-------|--------------------------|--------------------------|------------------------|
| Quantitative—non-randomized | | | |
| Aarons et al. 2009 [34] | Agency type | 1. organizational support for EBP --> provider attitudes towards EBP 2. 3 organizational support for EBP organizational support for EBP | 1,3 provider EBP use2. provider EBP attitudes |
| Becker et al. 2016 [35] | Training as usual, training + ongoing technical assistance, support from in-house champion, specialized training on change process, monthly conference calls and online forum to support change | Organizational readiness to change (motivation for change, adequacy of resources, staff attributes, organizational climate), perceived intervention characteristics (relative advantage, observability, trialability, compatibility, and complexity) | Adoption |
| Beenstock et al. 2012 [36] | Main place of work | Propensity to act | Referral of women to smoking cessation services |
| Beets et al. 2008 [37] | Perception of school climate | 1. Beliefs about responsibility to teach program 2. beliefs about responsibility to teach program --> attitudes towards program --> curriculum delivered | 1. Attitudes towards program 2. curriculum delivered to schoolwide material usage |
| Bonetti et al. 2009 [38] | Behavioral intention | Action planning | Placing fissure sealants |
| Chou et al. 2011 [39] | Receipt of individual performance feedback, clinician input into guideline implementation and quality improvement, clinician expectancy, clinician self-efficacy | Agreement with guidelines, adherence to guidelines, improved knowledge, practice delivery | Fidelity to screening patients for depression |
| Cummings et al. 2017 [40] | Culture, feedback, leadership and resources | Manager support, coaching conversations, job satisfaction | Conceptual research use, persuasive research use, instrumental research use |
| David and Schiff 2017 [41] | Child-parent psychotherapy social network Child-parent psychotherapy supervision | Self-efficacy | Number of child-parent psychotherapy cases, intention to use child-parent psychotherapy |
| Edmunds et al. 2014 [42] | Time following training | Time spent in consultation | Knowledge of cognitive behavioral therapy for anxiety, attitudes towards EBPs |
| Gnich et al. 2018 [43] | Pay-per item financial incentive | Knowledge, skills, social/professional role and identity, beliefs about consequences, motivation and goals (intention), environmental context and resources, social influences (norms), emotion, behavioral regulation | Fluoride varnish delivery |
| Guerrero et al. 2018 [44] | Top manager transformational leadership | Middle managers’ implementation leadership | Employee attitudes towards EBPs, EBP implementation |
| Huis et al. 2013 [45] | individual and organization targeted strategies (education, reminders, feedback), individual and organizational targeted strategies + team and leader strategy | Social influence, leadership, performance feedback | Handwashing fidelity |
| Little et al. 2015 [46] | Community priority, organizational support, program champion | beliefs about effectiveness of interventions --> funding to adopt program | Adoption |
| Llasus et al. 2014 [47] | EBP knowledge | Self confidence in one’s EBP competencies (defined as readiness) | EBP implementation behaviors |
| Nelson and Steele 2007 [48] | EBP training, openness of clinical setting to EBPs | Positive attitudes towards treatment research, negative attitudes towards treatment research | EBP use |
| Potthoff et al. 2017 [49] | Action planning, coping planning | Habit | Clinical behaviors (prescribing, advising, examining) |
| Presseau et al. 2016 [50] | Printed informational materials | Attitudes toward prescribing, subjective norms, perceived behavioral control, intention to prescribe | Self-reported prescribing behavior |
| Simmonds et al. 2012 [51] | Intolerance of uncertainty | Treatment orientation toward back pain | Recommendations to return to work 2. recommendations to return to usual activities,estimated risk of back pain |
Table 7 Model tested (Continued)

| Study                        | Independent variable (X)                      | Intervening variable (M)                       | Dependent variable (Y)                  |
|------------------------------|-----------------------------------------------|------------------------------------------------|-----------------------------------------|
| Stockdale et al. 2018 [52]   | Health care team communication                | Patient-provider communication                | Patient satisfaction with primary care provider |
| Wanless et al. 2015 [53]     | Use of responsive classroom practices, global emotional support, self-efficacy, collective responsibility | Teacher training engagement                  | Fidelity to intervention                |
| Yamada et al. 2017 [62]     | Instrumental research use, conceptual research use | Organizational context, leadership, culture, evaluation, social capital, informal interactions, formal interactions, resources, slack space, slack staff, slack time | Pain assessment, evidence-based pain procedure use, pain intensity |
| Mixed methods                |                                               | Evidence-based information in guideline, evidence-based information in screening module, discussions with peers, application of implementation tools, discussions with patients, lack of evidence about benefits, patients' screening expectations, fear of misdiagnosis, problems with having patient materials available | Use of breast cancer screening guidelines |
| Armson et al. 2018 [54]     | Implementation tools (printed education materials, informational video, decision aid) | Mediators: 1a. Performance reviews 1b. Human resources 1c. Training 1d. Funding 1e. Local social network involvement Moderator: 2/3. top manager support | 1, 2, 3 middle manager commitment to innovation |
| Birken et al. 2015 [55]     | 1. Top manager support                        |                                               |                                         |
|                             | 2. Performance reviews                        |                                               |                                         |
|                             | 3. Human resources                            |                                               |                                         |
| Kauth et al. 2010 [56]      | Facilitation + workshop, workshop             |                                               | % time conducting Cognitive Behavioral Therapy |
| Lukas et al. 2009 [57]      | Higher management support, group culture, hierarchical culture |                                               | Extent of implementation                |
| Panzano et al. 2012 [58]    | 1. Strategic fit of intervention              |                                               | 1. Fidelity to intervention 2. Assimilation |
|                             | 2. Climate for innovation                     |                                               |                                         |
| Rangachari et al. 2015 [59] | Emails containing intervention information and unit level adherence feedback + brief weekly training |                                               | Number of catheter days               |
| Shrubsole et al. 2018 [60]  | Tailored training intervention targeting information provision |                                               | Information provision/goal setting      |

Numbering is used to denote match variables across models; not all models tested the same sets of variables

(strategy) and adoption of the EBP (implementation outcome). Five studies included a mediation model in which X was an implementation strategy, Y was a clinical outcome, and M was an implementation outcome (strategy → implementation outcome → clinical outcome) [25, 26, 28, 29, 31].

Notable exceptions to model types

While the majority of quantitative studies tested a three-variable model, there were some notable exceptions. Several studies tested multiple three variable models that held the independent variable and mediator constant but tested the relation among several dependent variables. Several studies tested multiple three variable models that held the independent variable and dependent variables constant but tested several mediators.

Qualitative studies

Five studies included in this review utilized qualitative methods to explore potential mechanisms or mediators of change, though only one explicitly stated this goal in their aims [17]. Three studies utilized a comparative case study design incorporating a combination of interviews,
focus groups, observation, and document review, whereas two studies employed a cross-sectional descriptive design. Although three of the five studies reported their analytic design was informed by a theory or previously established model, only one study included an interview guide in which items were explicitly linked to theory [19]. All qualitative studies explored relations between multiple ecological levels, drawing connections between intra- and interpersonal behavioral constructs and organization or system level change.

Criteria for establishing mechanisms of change
Finally, with respect to the seven criteria for establishing mechanisms of change, the plausibility/coherence (i.e., a logical explanation of how the mechanism operates that incorporates relevant research findings) was the most frequently fulfilled requirement, met by 42 studies (91.3%). Although 20 studies (43.5%), of which 18 were quantitative, provided statistical evidence of a strong association between the dependent and independent variables, only 13 (28.2%) studies experimentally manipulated an implementation strategy or the proposed mediator or mechanism. Further, there was only one study that attempted to demonstrate a dose-response relation between mediators and outcomes. Most included studies (39; 84.8%) fulfilled three or fewer criteria, and only one study fulfilled six of the seven requirements for demonstrating a mechanism of change; see Table 8.

Discussion
Observations regarding mechanistic research in implementation science
Mechanism-focused implementation research is in an early phase of development, with only 46 studies identified in our systematic review across health disciplines broadly. Consistent with the field of implementation science, no single discipline is driving the conduct of mechanistic research, and a diverse array of methods (quantitative, qualitative, mixed methods) and designs (e.g., cross-sectional survey, longitudinal non-randomized, longitudinal randomized, etc.) have been used to examine mechanisms. Just over one-third of studies (N = 16; 34.8%) evaluated a mediation model with the implementation strategy as the independent variable, determinant as a putative mediator, and implementation outcome as the dependent variable. Although this was the most commonly reported model, we would expect a much higher proportion of studies testing mechanisms of implementation strategies given the ultimate goal of precise selection of strategies targeting key mechanisms of change. Studies sometimes evaluated models in which the determinant was the independent variable, another determinant was the putative mediator, and an implementation outcome was the dependent variable (N = 11; 23.9%). These models suggest an interest in understanding the cascading effect of changes in context on key outcomes, but without manipulating or evaluating an implementation strategy as the driver of observed change. Less common (only 5, 10.9%) were more complex models in which multiple mediators and outcomes and different levels of analyses were tested (e.g., [37, 39]), despite that this level of complexity is likely to characterize the reality of typical implementation contexts. Although there were several quantitative studies that did observe significant relations pointing toward a mediator, none met all criteria for establishing a mechanism.

Less than one-third of the studies experimentally manipulated the strategy-mechanism linkage. As the field progresses, we anticipate many more tests of this nature, which will allow us to discern how strategies exert their effect on outcomes of interest. However, implementation science will continue to be challenged by the costly nature of the type of experimental studies that would be needed to establish this type of evidence. Fortunately, methodological innovations that capitalize on recently funded implementation trials to engage in multilevel mediation modeling hold promise for the next iteration of mechanistic implementation research [14, 127] As this work unfolds, a number of scenarios are possible. For example, it is likely the case that multiple strategies can target the same mechanism; that a single strategy can target multiple mechanisms; and that mechanisms across multiple levels of analysis must be engaged for a given strategy to influence an outcome of interest. Accordingly, we expect great variability in model testing will continue and that more narrowly focused efforts will remain important contributions so long as shared conceptualization of mechanisms and related variables is embraced, articulated, and rigorously tested. As with other fields, we observed great variability in the degree to which mechanisms (and related variables of interest) were appropriately specified, operationalized, and measured. This misspecification coupled with the overall lack of high-quality studies (only three met 100% of the quality criteria), and the diversity in study methods, strategies tested, and mediating or moderating variables under consideration, we were unable to synthesize the findings across studies to point toward promising mechanisms.

The need for greater conceptual clarity and methodological advancements
Despite the important advances that the studies included in this review represent, there are clear conceptual and methodological issues that need to be addressed to allow future research to more systematically establish mechanisms. Table 1 offers a list of key terms and definitions for the field to consider. We suggest the term “mechanism” be used to reflect a process or event through which an implementation strategy operates to affect desired
Table 8 Kazdin criteria

|                          | Association | Specificity | Consistency | Manipulation | Timeline | Gradient | Plausibility | Total |
|--------------------------|-------------|-------------|-------------|--------------|----------|----------|--------------|-------|
| **Qualitative**          |             |             |             |              |          |          |              |       |
| Bardosh et al. 2017 [16] | 0           | 0           | 1           | 0            | 0        | 0        | 1            | 2     |
| Brewster et al. 2015 [17]| 0           | 0           | 1           | 0            | 0        | 0        | 1            | 2     |
| Carrera and Lambooij 2015 [18]| 0   | 0           | 0           | 0            | 0        | 0        | 1            | 1     |
| Frykman et al. 2014 [19] | 0           | 0           | 0           | 0            | 1        | 0        | 1            | 2     |
| Wiener-Ogilvie et al. 2008 [20]| 0 | 0           | 1           | 0            | 0        | 0        | 1            | 2     |
| **Quantitative—randomized** |           |             |             |              |          |          |              |       |
| Atkins et al. 2008 [21]  | 0           | 0           | 1           | 1            | 0        | 0        | 1            | 3     |
| Baer et al. 2009 [22]    | 0           | 0           | 1           | 0            | 1        | 0        | 1            | 3     |
| Bonetti et al. 2005 [23] | 1           | 1           | 1           | 0            | 1        | 0        | 1            | 5     |
| Garner et al. 2011 [24]  | 0           | 1           | 0           | 1            | 0        | 0        | 1            | 3     |
| Glisson et al. 2010 [25] | 0           | 0           | 0           | 1            | 1        | 0        | 1            | 3     |
| Holth et al. 2011 [26]   | 1           | 0           | 1           | 1            | 1        | 0        | 1            | 4     |
| Lee et al. 2018 [27]     | 0           | 0           | 0           | 0            | 0        | 0        | 1            | 1     |
| Lochman et al. 2009 [28] | 0           | 1           | 0           | 1            | 1        | 0        | 0            | 3     |
| Rapkin et al. 2017 [29]  | 1           | 0           | 0           | 0            | 1        | 1        | 1            | 4     |
| Rohrbach et al. 1993 [30]| 0           | 0           | 0           | 1            | 1        | 0        | 1            | 3     |
| Seys et al. 2018 [31]    | 1           | 1           | 0           | 1            | 1        | 0        | 1            | 5     |
| Williams et al. 2014 [32]| 0           | 1           | 0           | 1            | 1        | 0        | 1            | 4     |
| Williams et al. 2017 [33]| 1           | 1           | 1           | 1            | 1        | 0        | 1            | 6     |
| **Quantitative—non-randomized** |       |             |             |              |          |          |              |       |
| Aarons et al. 2009 [34]  | 1           | 0           | 1           | 0            | 0        | 0        | 1            | 4     |
| Becker et al. 2016 [35]  | 0           | 0           | 0           | 1            | 1        | 0        | 1            | 3     |
| Beenstock et al. 2012 [36]| 1         | 0           | 0           | 0            | 0        | 0        | 1            | 1     |
| Beets et al. 2008 [37]   | 1           | 0           | 1           | 0            | 0        | 0        | 1            | 3     |
| Bonetti et al. 2009 [38] | 1           | 0           | 1           | 0            | 0        | 0        | 1            | 3     |
| Chou et al. 2011 [39]    | 1           | 0           | 0           | 0            | 0        | 0        | 1            | 2     |
| Cummings et al. 2017 [40]| 0           | 0           | 1           | 0            | 0        | 0        | 1            | 3     |
| David and Schiff 2017 [41]| 1           | 0           | 1           | 0            | 0        | 0        | 1            | 3     |
| Edmunds et al. 2014 [42] | 0           | 0           | 0           | 0            | 1        | 0        | 0            | 1     |
| Gnich et al. 2018 [43]   | 0           | 0           | 1           | 0            | 1        | 0        | 1            | 3     |
| Guererro et al. 2018 [44]| 1           | 0           | 1           | 0            | 0        | 0        | 1            | 2     |
| Huis et al. 2013 [45]    | 0           | 0           | 0           | 1            | 1        | 0        | 1            | 3     |
| Little et al. 2015 [46]  | 1           | 0           | 0           | 0            | 0        | 0        | 1            | 2     |
| Llasus et al. 2014 [47]  | 1           | 0           | 1           | 0            | 0        | 0        | 1            | 3     |
| Nelson and Steele 2007 [48]| 1   | 0           | 0           | 0            | 0        | 0        | 1            | 2     |
| Potthoff et al. 2017 [49]| 1           | 0           | 0           | 0            | 0        | 0        | 1            | 2     |
| Pesseau et al. 2016 [50] | 0           | 0           | 1           | 0            | 0        | 0        | 1            | 2     |
| Simmonds et al. 2012 [51]| 1           | 0           | 0           | 0            | 0        | 0        | 1            | 2     |
| Stockdale et al. 2018 [52]| 1         | 0           | 0           | 0            | 0        | 0        | 1            | 2     |
| Wanless et al. 2015 [53] | 0           | 0           | 0           | 0            | 1        | 0        | 1            | 2     |
| **Mixed methods**        |             |             |             |              |          |          |              |       |
| Armson et al. 2018 [54]  | 0           | 0           | 1           | 0            | 0        | 0        | 1            | 2     |
| Birken et al. 2015 [55]  | 0           | 0           | 0           | 0            | 0        | 0        | 1            | 1     |
implementation outcomes. Consistent with existing criteria [4], mechanisms can only be confidently established via carefully designed (i.e., longitudinally; experimentally manipulated) empirical studies demonstrating a strong association, and ideally a dose-response relation, between an intervening variable and outcome (e.g., via qualitative data or mediation or moderator analyses) that are supported by very specific theoretical propositions observed consistently across multiple studies. We found the term “mediator” to be most frequently used in this systematic review, which can point toward a mechanism, but without consideration of these full criteria, detection of a mediator reflects a missed opportunity to contribute more meaningfully to the mechanisms literature.

Interestingly, the nearly half of studies (43.5%) treated a variable that many would conceptualize as a “determinant” as the independent variable in at least one proposed or tested mediation pathway. Presumably, if researchers are exploring the impact of a determinant on another determinant and then on an outcome, there must be a strategy (or action) that caused the change in the initial determinant. Or, it is possible that researchers are simply interested in the natural associations among these determinants to identify promising points of leverage. This is a prime example where the variable or overlapping use of concepts (i.e., calling all factors of interest “determinants”) becomes particularly problematic and undermines the capacity of the field to accumulate knowledge across studies in the service of establishing mechanisms. We contend that it is important to differentiate among concepts to use more meaningful terms like preconditions, putative mechanisms, proximal and distal outcomes, all of which were under-specified in the majority of the included studies. Several authors from our team have articulated an approach to building causal pathway diagrams [128] that clarifies that preconditions are necessary factors for a mechanism to be activated and proximal outcomes are the immediate result of a strategy that is realized only because the specific mechanism was activated. We conceptualize distal outcomes as the eight implementation outcomes articulated by Proctor and colleagues [129]. Disentangling these concepts can help characterize why strategies fail to exert an impact on an outcome of interest. Examples of each follow in the section below.

Conceptual and methodological recommendations for future research

Hypothesis generation

With greater precision among these concepts, the field can also generate and test more specific hypotheses about how and why key variables are related. This begins with laying out mechanistic research questions (e.g., How does a network intervention, like a learning collaborative, influence provider attitudes?) and generating theory-driven hypotheses. For instance, a testable hypothesis may be that learning collaboratives [strategy] operate through sharing [mechanism] of positive experiences with a new practice to influence provider attitudes [outcome]. As another example, clinical decision support [strategy] may act through helping the provider to remember [mechanism] to administer a screener [proximal outcome] and flagging this practice before an encounter may not allow the mechanism to be activated [precondition]. Finally, organizational strategy development [strategy] may have an effect because it means prioritizing competing demands [mechanism] to generate a positive implementation climate [proximal outcome]. Research questions that allow for specific mechanism-focused hypotheses have the potential to expedite the rate at which effective implementation strategies are identified.

Implementation theory

Ultimately, theory is necessary to drive hypotheses, explain implementation processes, and effectively inform implementation practice by providing guidance about when and in what contexts specific implementation strategies should or should not be used. Implementation theories can offer mechanisms that extend across levels of analysis (e.g., intrapersonal, interpersonal, organizational, community, macro policy [130]). However, there is a preponderance of frameworks and process models, with few theories in existence. Given that implementation is a process of behavior change at its core, in lieu of implementation-specific theories, many researchers draw upon classic theories from psychology, decision science, and organizational literatures, for
instance. Because of this, the majority of the identified studies explored intrapersonal-level mechanisms, driven by their testing of social psychological theories such as the theory of planned behavior [65] and social cognitive theory [76, 77, 99]. Nine studies cited the diffusion of innovations [63, 64] as a theory guiding their mechanism investigation, which does extend beyond intrapersonal to emphasize interpersonal, and to some degree community level mechanisms, although we did not see this materialize in the included study analyses [63–65, 76, 77]. Moving forward, developing and testing theory is critical for advancing the study of implementation mechanisms because theories (implicitly or explicitly) tend to identify putative mechanisms instead of immutable determinants.

Measurement
Inadequate measurement has the potential to undermine our ability to advance this area of research. Our coding indicated that mechanisms were assessed almost exclusively via self-report (questionnaire, interview, focus group) suggesting that researchers conceptualize the diverse array of mechanisms to be latent constructs and not directly observable. This may indeed be appropriate, given that mechanisms are typically processes like learning and reflecting that occur within an individual and it is their proximal outcomes that are directly observable (e.g., knowledge acquisition, confidence, perceived control). However, conceptual, theoretical, and empirical work is needed to (a) articulate the theorized mechanisms for the 70+ strategies and proximal outcomes [128], (b) identify measures of implementation mechanisms and evaluate their psychometric evidence base [131] and pragmatic qualities [132], and (c) attempt to identify and rate or develop objective measures of proximal outcomes for use in real-time experimental manipulations of mechanism-outcome pairings.

Quantitative analytic approaches
The multilevel interrelations of factors implicated in an implementation process also call for sophisticated quantitative and qualitative methods to uncover mechanisms. With respect to quantitative methods, it was surprising that the Baron and Kenny [78] approach to mediation testing remains most prevalent despite that most studies are statistically underpowered to use this approach, and the other most common approach (i.e., the Sobel test [79]) relies on an assumption that the sampling distribution of the mediation effect is normal [14, 133], neither of which were reported on in any of the 12 included studies that used these methods. Williams [14] suggests the product of coefficients approach [134, 135] is more appropriate for mediation analysis because it is a highly general approach to both single and multi-level mediation models that minimizes type I error rates, maximizes statistical power, and enhances accuracy of confidence intervals [14]. The application of moderated mediation models and mediated moderator models will allow for a nuanced understanding of the complex interrelations among factors implicated in an implementation process.

Qualitative analytic approaches
Because this was the first review of implementation mechanisms across health disciplines, we believed it was important to be inclusive with respect to methods employed. Qualitative studies are important to advancing research on implementation mechanisms in part because they offer a data collection method in lieu of having an established measure to assess mechanisms quantitatively. Qualitative research is important for informing measure development work, but also for theory development given the richness of the data that can be gleaned. Qualitative inquiry can be more directive by developing hypotheses and generating interview guides to directly test mechanisms. Diagramming and tracing causal linkages can be informed by qualitative inquiry in a structured way that is explicit with regard to how the data informs our understanding of mechanisms. This kind of directed qualitative research is called for in the United Kingdom’s MRC Guidance for Process Evaluation [136]. We encourage researchers internationally to adopt this approach as it would importantly advance us beyond the descriptive studies that currently dominate the field.

Limitations
There are several limitations to this study. First, we took an efficient approach to coding for study quality when applying the MMAT. Although it was a strength that we evaluated study quality, the majority of studies were assessed only by one research specialist. Second, we may have overlooked relevant process evaluations conducted in the UK where MRC Guidance stipulates inclusion of mechanisms that may have been described using terms not included in our search string. Third, although we identified several realist reviews, we did not include them in our systematic review because they conceptualize mechanisms differently than how they are treated in this review [137]. That is, realist synthesis posits that interventions are theories and that they imply specific mechanisms of action instead of separating mechanisms from the implementation strategies/interventions themselves [138]. Thus, including the realist operationalization would have further confused an already disharmonized literature with respect to mechanisms terminology but ultimately synthesizing findings from realist reviews with standard implementation mechanism evaluations will be important. Fourth, our characterization of the models tested in the identified studies may not reflect those intended by researchers
given our attempt to offer conceptual consistency across studies, although we did reach out to corresponding authors for whom we wished to seek clarification on their study. Finally, because of the diversity of study designs and methods, and the inconsistent use of relevant terms, we are unable to synthesize across the studies and report on any robustly established mechanisms.

Conclusion
This study represents the first systematic review of implementation mechanisms in health. Our inclusive approach yielded 46 qualitative, quantitative, and mixed methods studies, none of which met all seven criteria (i.e., strong association, specificity, consistency, experimental manipulation, timeline, gradient, plausibility or coherence) that are deemed critical for empirically establishing mechanisms. We found nine unique versions of models that attempted to uncover mechanisms, with only six exploring mediators of implementation strategies. The results of this review indicated inconsistent use of relevant terms (e.g., mechanisms, determinants) for which we offer guidance to achieve precision and encourage greater specificity in articulating research questions and hypotheses that allow for careful testing of causal relations among variables of interest. Implementation science will benefit from both quantitative and qualitative research that is more explicit in their attempt to uncover mechanisms. In doing so, our research will allow us to test the idea that more is better and move toward parsimony both for standardized and tailored approaches to implementation.

Supplementary information
Supplementary information accompanies this paper at https://doi.org/10.1186/s13012-020-00983-3.

Additional file 1: Figure S1. Inclusion and Exclusion Criteria and Definitions.
Additional file 2. PRISMA 2009 Checklist.
Additional file 3. Emergent Mechanism Models.

Acknowledgments
Not applicable.

Availability of data and material
The authors are willing to share the raw data tables that informed the summary tables included in this manuscript.

Authors’ contributions
CCL conceptualized the larger study and articulated the research questions with all coauthors. CCL, MRB, and CWB designed the approach with feedback from all coauthors. MRB and CWB executed the systematic search with oversight and checking by CCL. MRB led the data extraction and CWB led the study appraisal. All authors contributed to the discussion and reviewed and approved the manuscript.

Funding
This project was supported by grant number R13HS025632 from the Agency for Healthcare Research and Quality. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
1Kaiser Permanente Washington Health Research Institute, 1730 Minor Avenue, Suite 1600, Seattle, WA 98101, USA. 2Department of Psychological and Brain Sciences, Indiana University, 1101 E 10th Street, Bloomington, IN 47405, USA. 3Department of Psychiatry and Behavioral Sciences, School of Medicine, University of Washington, 1959 NE Pacific Avenue, Seattle, WA 98195, USA. 4Department of Psychology, University of California Los Angeles, 1177 Franz Hall, 502 Portola Plaza, Los Angeles, CA 90095, USA. 5Brown School, Washington University in St. Louis, 1 Brookings Drive, St. Louis, MO 63130, USA. 6Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, 3535 Market Street, Philadelphia, PA 19104, USA. 7Department of Research and Evaluation, Kaiser Permanente Southern California, 100 S Los Robles Avenue, Pasadena, CA 91101, USA. 8Department of Psychiatry, School of Medicine, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA. 9Department of Health Services, University of Washington, 1959 NE Pacific Street, Seattle, WA 98195, USA. 10Division of Cancer Control and Population Science, National Cancer Institute, 9609 Medical Center Drive, Rockville, MD 20850, USA.

Received: 12 November 2019 Accepted: 12 March 2020
Published online: 16 April 2020

References
1. Eccles MP, Mittman BS. Welcome to implementation science. Implement Sci. 2006;1(1):1.
2. Powell BJ, McMillen JC, Proctor EK, Carpenter CR, Griffee RT, Bunger AC, et al. A compilation of strategies for implementing clinical innovations in health and mental health. Med Care Res Rev. 2012;69(2):123–57.
3. Powell BJ, Waltz TJ, Chinnman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implement Sci. 2015;10:21.
4. Kazdin AE. Mediators and mechanisms of change in psychotherapy research. Annu Rev Clin Psychol. 2007;3(1):1–27.
5. Kraemer HC, Wilson GT, Fairburn CG, Agras WS. Mediators and moderators of treatment effects in randomized clinical trials. Arch Gen Psychiatry. 2002; 59(10):877–83.
6. Gerring J. Social science methodology: a criterial framework. Cambridge: Cambridge University Press; 2001.
7. Frazier PA, Tix AP, Barron KE. Testing moderator and mediator effects in counseling psychology research. US: American Psychological Association; 2004. p. 115–34.
8. Hill AB. The Environment and Disease: Association or Causation? Proc R Soc Med. 1965;58:205–30.
9. Bosch M, van der Weijden T, Wensing M, Grof R. Tailoring quality improvement interventions to identified barriers: a multiple case analysis. J Eval Clin Pract. 2007;13(2):161–8.
10. Claridge JA, Fabian TC. History and development of evidence-based medicine. World J Surg. 2005;29(5):547–53.
11. Cook SC, Schwartz AC, Kaslow NJ. Evidence-Based psychotherapy: advantages and challenges. Neurotherapeutics. 2017;14(3):537–45.
12. Dissemination and Implementation Research in Health (R01 Clinical Trial Optional). National Institutes of Health (NIH); 2019. https://grants.nih.gov/grants/guide/ pa-files/PAR-19-274.html.
115. Bentler PM. On tests and indices for evaluating structural models. Personal Individ Differ. 2007;42(5):825–9.

116. Mackinnon DP, Fairchild AJ, Fritz MS. Mediation analysis. Annu Rev Psychol. 2007;58:593–614.

117. Graham J, Logan J, Harrison M, Straus S, Tetroe J, Caswell W, et al. Lost in knowledge translation: time for a map? J Contin Educ Health Prof. 2006;26.

118. Epstein S. Cognitive-experiential self-theory. In: Pervin LA, editor. Handbook of personality: theory and research. New York: Guilford; 1990. p. 165–92.

119. Karlson KB, Holm A, Breen R. Comparing Regression coefficients between same-sample Nested models using logit and probit: a new method. Sociological Methodology. 2012;42(1):274–301.

120. Rycroft-Malone J, Kitson A, Harvey G, McCormack B, Seers K, Titchen A, et al. Ingredients for change: revisiting a conceptual framework. BMJ Qual Saf. 2002;11(2):174–80.

121. Yuki G, Gordon A, Taber T. A hierarchical taxonomy of leadership behavior: integrating a half century of behavior research. J Leadersh Organ Stud. 2002;9(1):15–32.

122. Shrout PE, Bolger N. Mediation in experimental and nonexperimental studies: new procedures and recommendations. Psychol Methods. 2002;7(4):422–45.

123. Fixsen DL, Naoom SF, Blase KA, Friedman RM. Implementation research: a synthesis of the literature; 2005.

124. Frambach R. An integrated model of organizational adoption and diffusion of innovations. Eur J Mark. 1993;27(3):22–41.

125. Institute of Medicine (IOM). Crossing the quality chasm: a new health system for the 21st century. Washington, DC: Institute of Medicine, National Academy Press; 2001.

126. Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behav Res Methods. 2008;40(3):879–91.

127. Stahmer AC, Suheineich J, Schetter PL, McGee HE. Exploring multi-level system factors facilitating educator training and implementation of evidence-based practices (EBP), a study protocol. Implement Sci. 2018;13(1):3.

128. Lewis CC, Klasnja P, Powell BJ, Lyon AR, Tuzzio L, Jones S, et al. From classification to causality: advancing understanding of mechanisms of change in implementation science. Front Public Health. 2018;6:136.

129. Proctor E, Grimmer-SprBMJ : British Medical Journal. 2015;350:h1258.

130. Powell BJ, Stanick CF, Halko HM, Dorsey CN, Weiner BJ, Barwick MA, et al. Toward criteria for pragmatic measurement in implementation research and practice: a stakeholder-driven approach using concept mapping. Implement Sci. 2017;12(1):118.

131. Wu AD, Zumbo BD. Understanding and using mediators and moderators. Soc Indic Res. 2007;87(3):367.

132. MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. A comparison of methods to test mediation and other intervening variable effects. Psychol Methods. 2002;7(1):83.

133. Pituch KA, Murphy DL, Tate RL. Three-level models for indirect effects in school- and class-randomized experiments in education. J Exp Educ. 2009;78(1):60–95.

134. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ : British Medical Journal. 2015;350:h1258.

135. Pawson R, Manzano-Santaella A. A realist diagnostic workshop. Evaluation. 2012;19(2):76–91.

136. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ : British Medical Journal. 2015;350:h1258.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.