



ELSEVIER

Contents lists available at ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

Quantitative approaches for the evaluation of implementation research studies

Justin D. Smith^{a,*}, Mohamed Hasan^b

^a Center for Prevention Implementation Methodology (Ce-PIM) for Drug Abuse and HIV, Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, 750 N Lake Shore Dr., Chicago, IL, USA

^b Center for Healthcare Studies, Institute of Public Health and Medicine, Northwestern University Feinberg School of Medicine, 633 N St. Claire St., Chicago, Illinois, USA

ARTICLE INFO

Keywords:

Implementation measurement
Implementation research
Summative evaluation

ABSTRACT

Implementation research necessitates a shift from clinical trial methods in both the conduct of the study and in the way that it is evaluated given the focus on the impact of implementation strategies. That is, the methods or techniques to support the adoption and delivery of a clinical or preventive intervention, program, or policy. As strategies target one or more levels within the service delivery system, evaluating their impact needs to follow suit. This article discusses the methods and practices involved in quantitative evaluations of implementation research studies. We focus on evaluation methods that characterize and quantify the overall impacts of an implementation strategy on various outcomes. This article discusses available measurement methods for common quantitative implementation outcomes involved in such an evaluation—adoption, fidelity, implementation cost, reach, and sustainment—and the sources of such data for these metrics using established taxonomies and frameworks. Last, we present an example of a quantitative evaluation from an ongoing randomized rollout implementation trial of the Collaborative Care Model for depression management in a large primary healthcare system.

1. Background

As part of this special issue on implementation science, this article discusses quantitative methods for evaluating implementation research studies and presents an example of an ongoing implementation trial for illustrative purposes. We focus on what is called “summative evaluation,” which characterizes and quantifies the impacts of an implementation strategy on various outcomes (Gaglio and Glasgow, 2017). This type of evaluation involves aggregation methods conducted at the end of a study to assess the success of an implementation strategy on the adoption, delivery, and sustainment of an evidence-based practice (EBP), and the cost associated with implementation (Bauer et al., 2015). These results help decision makers understand the overall worth of an implementation strategy and whether to upscale, modify, or discontinue (Bauer et al., 2015). This topic complements others in this issue on formative evaluation (Elwy & Finley, *this issue*) and qualitative methods (Hamilton and Finley, 2019), which are also used in implementation research evaluation.

Implementation research, as defined by the United States National Institutes of Health (NIH), is “the scientific study of the use of strategies

[italics added] to adopt and integrate evidence-based health interventions into clinical and community settings in order to improve patient outcomes and benefit population health. Implementation research seeks to understand the behavior of healthcare professionals and support staff, healthcare organizations, healthcare consumers and family members, and policymakers in context as key influences on the adoption, implementation and sustainability of evidence-based interventions and guidelines” (Department of Health and Human Services, 2019). Implementation strategies are methods or techniques used to enhance the adoption, implementation, and sustainability of a clinical program or practice (Powell et al., 2015).

To grasp the evaluation methods used in implementation research, one must appreciate the nature of this research and how the study designs, aims, and measures differ in fundamental ways from those methods with which readers will be most familiar—that is, evaluations of clinical efficacy or effectiveness trials. First, whereas clinical intervention research focuses on how a given clinical intervention—meaning a pill, program, practice, principle, product, policy, or procedure (Brown et al., 2017)—affects a health outcome at the patient level, implementation research focuses on how systems can take that

Abbreviations: CCM, collaborative care model; EHR, electronic health record

* Corresponding author.

E-mail addresses: jd.smith@northwestern.edu (J.D. Smith), mohamedhasan2020@u.northwestern.edu (M. Hasan).

<https://doi.org/10.1016/j.psychres.2019.112521>

Received 30 March 2019; Received in revised form 14 August 2019; Accepted 16 August 2019

0165-1781/ © 2019 Elsevier B.V. All rights reserved.

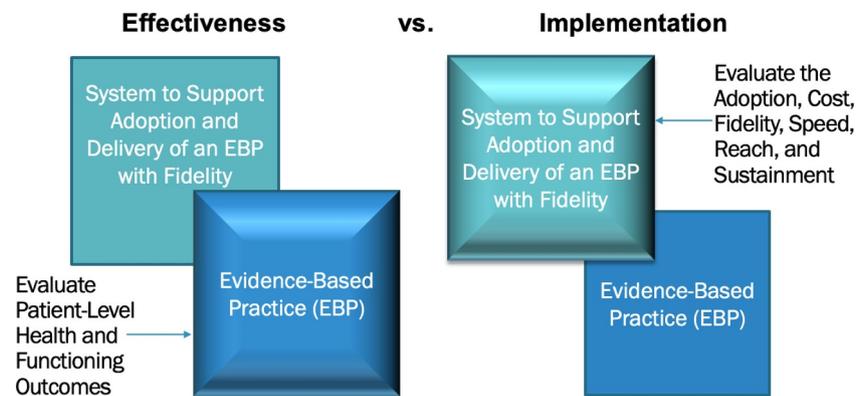


Fig. 1. Emphasis and outcomes evaluated in clinical effectiveness versus implementation research.
Note. Adapted from a slide developed by C. Hendricks Brown.

intervention to scale in order to improve health outcomes of the broader community (Colditz and Emmons, 2017). Thus, when implementation strategies are the focus, the outcomes evaluated are at the system level. Fig. 1 illustrates the emphasis (foreground box) of effectiveness versus implementation research and the corresponding outcomes that would be included in the evaluation. This difference can be illustrated by “hybrid trials” in which effectiveness and implementation are evaluated simultaneously but with different outcomes for each aim (Curran et al., 2012; also see Landes et al., *this issue*).

2. Design considerations for evaluating implementation research studies

The stark contrast between the emphasis in implementation versus effectiveness trials occurs largely because implementation strategies most often, but not always, target one or more levels within the system that supports the adoption and implementation of the intervention, such as the provider, clinic, school, health department, or even state or national levels (Powell et al., 2015). Implementation strategies are discussed in this issue by Kirchner et al. (2019). With the focus on levels within which patients who receive the clinical or preventive intervention are embedded, research designs in implementation research follow suit. The choice of a study design to evaluate an implementation strategy influences the confidence in the association drawn between a strategy and an observed effect (Grimshaw et al., 2000). Strong designs and methodologically-robust studies support the validity of the evaluations and provide evidence likely to be used by policy makers. Study designs are generally classified into observational (descriptive) and experimental/quasi-experimental.

Brown et al. (2017) described three broad types of designs for implementation research. (1) *Within-site designs* involve evaluation of the effects of implementation strategies within a single service system unit (e.g., clinic, hospital). Common within-site designs include post, pre-post, and interrupted time series. While these designs are simple and can be useful for understanding the impact in a local context (Cheung and Duan, 2014), they contribute limited generalizable knowledge due to the biases inherent small-sample studies with no direct comparison condition. Brown et al. describe two broad design types can be used to create generalizable knowledge as they inherently involve multiple units for aggregation and comparison using the evaluation methods described in this article. (2) *Between-site designs* involve comparison of outcomes between two or more service system units or clusters/groups of units. While they commonly involve the testing of a novel implementation strategy compared to routine practice (i.e., implementation as usual), they can also be head-to-head tests of two or more novel implementation strategies for the same intervention, which we refer to as a comparative implementation trial (e.g., Smith et al., 2019). (3) *Within- and between-site designs* add a time-based crossover

for each unit in which they begin in one condition—usually routine practice—and then move to a second condition involving the introduction of the implementation strategy. We refer to this category as rollout trials, which includes the stepped-wedge and dynamic wait-list design (Brown et al., 2017; Landsverk et al., 2017; Wyman et al., 2015). Designs for implementation research are discussed in this issue by Miller and colleagues.

3. Quantitative methods for evaluating implementation outcomes

While summative evaluation is distinguishable from formative evaluation (see Elwy & Finley, *this issue*), proper understanding of the implementation strategy requires using both methods, perhaps at different stages of implementation research (The Health Foundation, 2015). Formative evaluation is a rigorous assessment process designed to identify potential and actual influences on the effectiveness of implementation efforts (Stetler et al., 2006). Earlier stages of implementation research might rely solely on formative evaluation and the use of qualitative and mixed methods approaches. In contrast, later stage implementation research involves powered tests of the effect of one or more implementation strategies and are thus likely to use a between-site or a within- and between-site research design with at least one quantitative outcome. Quantitative methods are especially important to explore the extent and variation of change (within and across units) induced by the implementation strategies.

Proctor et al. (2011) provide a taxonomy of available implementation outcomes, which include acceptability, adoption, appropriateness, feasibility, fidelity, implementation cost, penetration/reach, and sustainability/sustainment. Table 1 in this article presents a modified version of Table 1 from Proctor et al. (2011), focusing only on the quantitative measurement characteristics of these outcomes. Table 1 also includes the additional metrics of speed and quantity, which will be discussed in more detail in the case example. As noted in Table 1, and by Proctor et al. (2011), certain outcomes are more applicable at earlier versus later stages of implementation research. A recent review of implementation research in the field of HIV indicated that earlier stage implementation research was more likely to focus on acceptability and feasibility, whereas later stage testing of implementation strategies focused less on these and more on adoption, cost, penetration/reach, fidelity, and sustainability (Smith et al., 2019). These sources of quantitative information are at multiple levels in the service delivery system, such as the intervention delivery agent, leadership, and key stakeholders in and outside of a particular delivery system (Brown et al., 2013).

Methods for quantitative data collection include structured surveys; use of administrative records, including payor and health expenditure records; extraction from the electronic health record (EHR); and direct observation. Structured surveys are commonly used to assess attitudes

Table 1
Quantitative measurement characteristics of common implementation outcomes.

Implementation outcome	Level of analysis	Other terms in the literature	Salience by implementation stage	Quantitative measurement method	Example from the published literature
Acceptability	Individual provider Individual consumer	Satisfaction with various aspects of the innovation (e.g. content, complexity, comfort, delivery, and credibility)	Early for adoption Ongoing for penetration Late for sustainability	Survey Administrative data Refused/blank	Auslander, W., McGinnis, H., Tlapak, S., Smith, P., Foster, A., Edmond, T., & Dunn, J. (2017). Adaptation and implementation of a trauma-focused cognitive behavioral intervention for girls in child welfare. <i>The American Journal of Orthopsychiatry</i> , 87(3), 206–215. doi:10.1037/ort0000233 https://www.ncbi.nlm.nih.gov/pubmed/27977284
Adoption	Individual provider Organization or setting	Uptake; utilization; initial implementation; intention to try	Early to mid	Administrative data Observation Survey	Knudsen, H. K., & Roman, P. M. (2014). Dissemination, adoption, and implementation of acamprostate for treating alcohol use disorders. <i>Journal of studies on alcohol and drugs</i> , 75(3), 467–475. doi:10.15288/jsad.2014.75.467 https://www.ncbi.nlm.nih.gov/pubmed/24766759
Appropriateness	Individual provider Individual consumer Organization or setting	Perceived fit; relevance; compatibility; suitability; usefulness; practicability	Early (prior to adoption)	Survey	Proctor, E., Ramsey, A. T., Brown, M. T., Malone, S., Hooley, C., & McKay, V. (2019). Training in Implementation Practice Leadership (TRIPLE): evaluation of a novel practice change strategy in behavioral health organizations. <i>Implementation science : IS</i> , 14(1), 66. doi:10.1186/s13012-019-0906-2 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6585005/
Feasibility	Individual providers Organization or setting	Actual fit or utility; suitability for everyday use; practicability	Early (during adoption)	Survey Administrative data	Lyon, A. R., Bruns, E. J., Ludwig, K., Stoep, A. V., Pullmann, M. D., Dorsey, S., ... McCauley, E. (2015). The Brief Intervention for School Clinicians (BRISC): A mixed-methods evaluation of feasibility, acceptability, and contextual appropriateness. <i>School mental health</i> , 7(4), 273–286. doi:10.1007/s12310-015-9153-0 https://www.ncbi.nlm.nih.gov/pubmed/26688700
Fidelity	Individual provider Program	Delivered as intended; adherence; integrity; quality of program delivery	Early to mid	Observation Checklists Self-report	Smith, J. D., Dishion, T. J., Shaw, D. S., & Wilson, M. N. (2013). Indirect effects of fidelity to the family check-up on changes in parenting and early childhood problem behaviors. <i>Journal of consulting and clinical psychology</i> , 81(6), 962–974. doi:10.1037/a0033950 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3852198/
Implementation cost	Provider or providing Institution Payor Individual consumer	Marginal cost; cost-effectiveness; cost-benefit	Early for adoption and feasibility Mid for penetration Late for sustainability	Administrative data	Jordan N, Graham AK, Berkel C, Smith JD (2019). Budget impact analysis of preparing to implement the Family Check-Up 4 Health in primary care to reduce pediatric obesity. <i>Prevention science</i> , 20(5), 655–664. doi:10.1007/s11121-018-0970-x https://www.ncbi.nlm.nih.gov/pubmed/30613852
Penetration/Reach	Organization or setting	Level of institutionalization? Spread? Service access?	Mid to late	Case audit Checklists	Emily M. Woltmann, M. S. W., Rob Whitley, P. D., Gregory J. McHugo, P. D., Mary Brunette, M. D., William C. Torrey, M. D., Laura Coots, M. S., ... Robert E. Drake, M. D., Ph.D., (2008). The Role of Staff Turnover in the Implementation of Evidence-Based Practices in Mental Health Care. <i>Psychiatric Services</i> , 59(7), 732–737. doi:10.1176/ps.2008.59.7.732 https://www.ncbi.nlm.nih.gov/pubmed/18586989
Sustainability	Administrators Organization or setting	Maintenance; continuation; durability; incorporation; integration; sustained use; institutionalization; routinization;	Late	Case audit Checklists Survey	Scudder, A. T., Taber-Thomas, S. M., Schaffner, K., Pemberton, J. R., Hunter, L., & Herschell, A. D. (2017). A mixed-methods study of system-level sustainability of evidence-based practices in 12 large-scale implementation initiatives. <i>Health research policy and systems</i> , 15(1), 102. doi:10.1186/s12961-017-0230-8 https://www.ncbi.nlm.nih.gov/pubmed/29216886
Quantity	Organization or setting	Proportion; quantity	Early through late	Administrative data Observation	Brown, C. H., Chamberlain, P., Saldana, L., Padgett, C., Wang, W., & Cruden, G. (2014). Evaluation of two implementation strategies in 51 child county public service systems in two states: results of a cluster randomized head-to-head implementation trial. <i>Implementation science</i> , 9, 134. doi:10.1186/s13012-014-0134-8 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4201704/

(continued on next page)

Table 1 (continued)

Implementation outcome	Level of analysis	Other terms in the literature	Salience by implementation stage	Quantitative measurement method	Example from the published literature
Speed	Organization or setting	Duration (speed)	Early through late	Administrative data Observation	Saldana, L., Chamberlain, P., Wang, W., & Hendricks Brown, C. (2012). Predicting program start-up using the stages of implementation measure. <i>Administration and policy in mental health</i> , 39(6), 419–425. doi:10.1007/s10488-011-0363-y https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3212640/

Note. This table is modeled after Table 1 in the Proctor et al. (2011) article.

and perceptions of providers and patients concerning such factors as the ability to sustain the intervention and a host of potential facilitators and barriers to implementation (e.g., Bertrand et al., 2009; Luke et al., 2014). Administrative databases and the EHR are used to assess aspects of intervention delivery that result from the implementation strategies (Bauer et al., 2015). Although the EHR supports automatic and cumulative data acquisition, its utility for measuring implementation outcomes is limited depending on the type of implementation strategy and the intervention. For example, it is well suited for gathering data on EHR-based implementation strategies, such as clinical decision supports and symptom screening, but less useful for behaviors that would not otherwise be documented in the EHR (e.g., effects of a learning collaborative on adoption of a cognitive behavioral therapy protocol). Last, observational assessment of implementation is fairly common but resource intensive, which limits its use outside of funded research. This is particularly germane to assessing fidelity of implementation, which is commonly observational in funded research but is rarely done when the intervention is adopted under real-world circumstances (Schoenwald et al., 2011). The costs associated with observational fidelity measurement has led to promising efforts to automate this process with machine learning methods (e.g., Imel et al., 2019).

Quantitative evaluation of implementation research studies most commonly involves assessment of multiple outcome metrics to garner a comprehensive appraisal of the effects of the implementation strategy. This is due in large part to the interrelatedness and interdependence of these metrics. A shortcoming of the Proctor et al. (2011) taxonomy is that it does not specify relations between outcomes, rather they are simply listed. The RE-AIM evaluation framework (Gaglio et al., 2013; Glasgow et al., 1999) is commonly used and includes consideration of the interrelatedness between both the implementation outcomes and the clinical effectiveness of the intervention being implemented. Thus, it is particularly well-suited for effectiveness-implementation hybrid trials (Curran et al., 2012; also see Landes et al., *this issue*). RE-AIM stands for Reach, Effectiveness (of the clinical or preventive intervention), Adoption, Implementation, and Maintenance. Each metric is important for determining the overall public health impact of the implementation, but they are somewhat interdependent. As such, RE-AIM dimensions can be presented in some combination, such as the “public health impact” metric (reach rate multiplied by the effect size of the intervention) (Glasgow et al., 2006). RE-AIM is one in a class of evaluation frameworks. For a review, see Tabak et al. (2012).

4. Resources for quantitative evaluation in implementation research

There are a number of useful resources for the quantitative measures used to evaluate implementation research studies. First is the Instrument Review Project affiliated with the Society for Implementation Research Collaboration (Lewis et al., 2015). The results of this systematic review of measures indicated significant variability in the coverage of measures across implementation outcomes and salient determinants of implementation (commonly referred to as barriers and facilitators). The authors reviewed each identified measure for the psychometric properties of internal consistency, structural validity, predictive validity, having norms, responsiveness, and usability (pragmatism). Few measures were deemed high-quality and psychometrically sound due in large part to not using gold-standard measure development methods. This review is ongoing and a website (<https://societyforimplementationresearchcollaboration.org/sirc-instrument-project/>) is continuously updated to reflect completed work, as well as emerging measures in the field, and is available to members of the society. A number of articles and book chapters provide critical discussions of the state of measurement in implementation research, noting the need for validation of instruments, use across studies, and pragmatism (Emmons et al., 2012; Lewis et al., 2015, 2017; Martinez et al., 2014; Rabin et al., 2016).

The RE-AIM website also includes various means of operationalizing the components of this evaluation framework (<http://www.re-aim.org/resources-and-tools/measures-and-checklists/>) and recent reviews of the use of RE-AIM are also helpful when planning a quantitative evaluation (Gaglio et al., 2013; Glasgow et al., 2019). Additionally, the Grid-Enabled Measures Database (GEM), hosted by the National Cancer Institute, has an ever-growing list of implementation-related measures (130 as of July, 2019) with a general rating by users (<https://www.gem-measures.org/public/wsmeasures.aspx?cat=8&aid=1&wid=11>). Last, Rabin et al. (2016) provide an environmental scan of resources for measures in implementation and dissemination science.

5. Pragmatism: reducing measurement burden

An emphasis in the field has been on finding ways to reduce the measurement burden on implementers, and to a lesser extent on implementation researchers to reduce costs and increase the pace of dissemination (Glasgow et al., 2019; Glasgow and Riley, 2013). Powell et al. (2017) established criteria for pragmatic measures that resulted in four distinct categories: (1) acceptable, (2) compatible, (3) easy, and (4) useful; next steps are to develop consensus regarding the most important criteria and developing quantifiable rating criteria for assessing implementation measures on their pragmatism. Advancements have occurred using technology for the evaluation of implementation (Brown et al., 2015). For example, automated and unobtrusive implementation measures can greatly reduce stakeholder burden and increase response rates. As an example, our group (Wang et al., 2016) conducted a proof-of-concept demonstrating the use text analysis to automatically classify the completion of implementation activities using communication logs between implementer and implementing agency. As mentioned earlier in this article, researchers have begun to automate the assessment of implementation fidelity to such evidence-based interventions as motivational interviewing (e.g., Imel et al., 2019; Xiao et al., 2015), and this work is expanding to other intervention protocols to aid in implementation quality (Smith et al., 2018).

6. Example of a quantitative evaluation of an implementation research study

We now present the quantitative evaluation plan for an ongoing hybrid type II effectiveness-implementation trial (see Landes et al., *this issue*) examining the effectiveness and implementation of the Collaborative Care Model (CCM; Unützer et al., 2002) for the management of depression in adult primary care clinics of Northwestern Medicine (Principal Investigator: Smith). CCM is a structure for population-based management of depression involving the primary care provider, a behavioral care manager, and a consulting psychiatrist. A meta-analysis of 79 randomized trials ($n = 24,308$), concluded that CCM is more effective than standard care for short- and long-term treatment of depression (Archer et al., 2012). CCM has also been shown to provide good economic value (Jacob et al., 2012).

Our study involves 11 primary care practices in a rollout implementation design (see Fig. 2). Randomization in roll-out designs occurs by start time of the implementation strategy, and ensures confidence in the results of the evaluation because known and unknown biases are equally distributed in the case and control groups (Grimshaw et al., 2000). Rollout trials are both powerful and practical as many organizations feel it is unethical to withhold effective interventions, and roll-out designs reduce the logistic and resource demands of delivering the strategy to all units simultaneously. The *co-primary aims* of the study concern the effectiveness of CCM and its implementation, respectively: 1) Test the effectiveness of CCM to improve depression symptomatology and access to psychiatric services within the primary care environment; and 2) Evaluate the impact of our strategy package on the progressive improvement in speed and quantity

of CCM implementation over successive clinics. We will use training and educational implementation strategies, provided to primary care providers, support staff (e.g., nurses, medical assistants), and to practice and system leadership, as well as monitoring and feedback to the practices. Fig. 3 summarizes the quantitative evaluation being conducted in this trial using the RE-AIM framework.

6.1. EHR and other administrative data sources

As this is a Type 2 effectiveness-implementation hybrid trial, Aim 1 encompasses both *reach*—an implementation outcome—of depression management by CCM within primary care—and the *effectiveness* of CCM at improving patient and service outcomes. Within RE-AIM, the Public Health Impact metric is effectiveness (effect size) multiplied by reach rate. EHR and administrative data are being used to evaluate the primary implementation outcomes of reach (i.e., the proportion of patients in the practice who are eligible for CCM and who are referred). The reach rates achieved after implementation of CCM can be compared to rates of mental health contact for patients with depression prior to implementation as well as to that achieved by other CCM implementation evaluations in the literature.

The primary effectiveness outcome of CCM is the reduction of patients' depression symptom severity. De-identified longitudinal patient outcome data from the EHR—principally depression diagnosis and scores on the PHQ-9 (Kroenke et al., 2001)—will be analyzed to evaluate the impact of CCM. Other indicators of the effectiveness of CCM will be evaluated as well but are not discussed here as they are likely to be familiar to most readers with knowledge of clinical trials. Service outcomes, from the Institute of Medicine's Standards of Care (Institute of Medicine Committee on Crossing the Quality Chasm, 2006), centered on providing care that is effective (providing services based on scientific knowledge to all who could benefit and refraining from providing services to those not likely to benefit), timely (reducing waits and sometimes harmful delays for both those who receive and those who give care), and equitable (providing care that does not vary in quality because of personal characteristics such as gender, ethnicity, geographic location, and socioeconomic status). We also sought to provide care that is safe, patient-centered, and efficient.

EHR data will also be used to determine *adoption* of CCM (i.e., the number of providers with eligible patients who refer to CCM). This can be accomplished by tracking patient screening results and intakes completed by the CCM behavioral care manager within the primary care clinician's encounter record.

6.2. Speed and quantity of implementation

Achievement of Aim 2 requires an evaluation approach and an appropriate trial design to obtain results that can contribute to generalizable knowledge. A rigorous rollout implementation trial design, with matched-pair randomization to when the practice would change from usual care to CCM was devised. Fig. 2 provides a schematic of the design with the timing of the crossover from standard practice to CCM implementation. The first thing one will notice about the design is that the sequential nature of the rollout in which implementation at earlier sites precedes the onset of implementation in later sites. This suggests the potential to learn from successes and challenges to improve implementation efficiency (speed) over time. We will use the Universal SIC® (Saldana et al., 2015), a date-based, observational measure, to capture the speed of implementation of various activities needed to successfully implement CCM, such as “establishing a workflow”, “preparing for training”, and “behavioral care manager hired.” This measure is completed by practice staff and members of the implementation team based on their direct knowledge of precisely when the activity was completed. Using the completion date of each activity, we will analyze the time elapsed in each practice to complete each stage (Duration Score). Then, we will calculate the percentage of stages completed

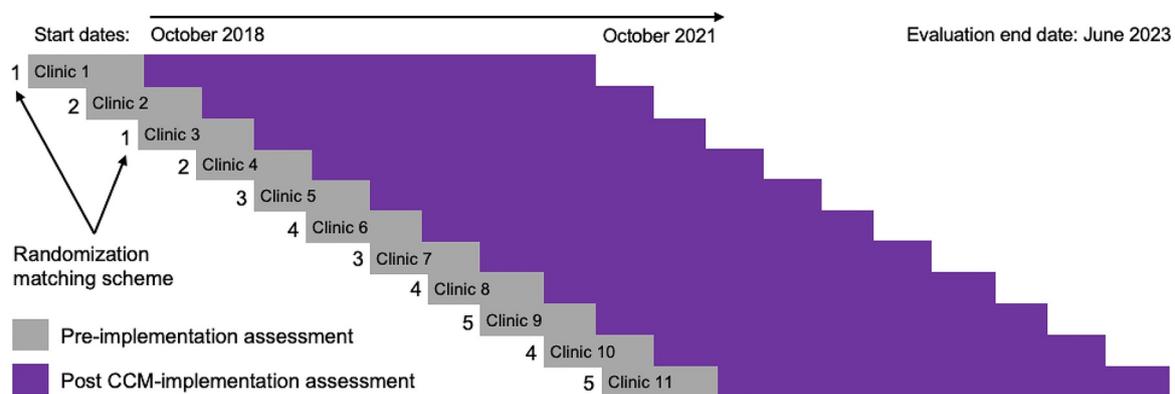


Fig. 2. Design and timeline of randomized rollout implementation trial of CCM.

Note. CCM = Collaborative Care Model. Clinics will have a staggered start every 3–4 months randomized using a matching scheme. Pre-implementation assessment period is 4 months. Evaluation of CCM implementation will be a minimum of 24 months at each clinic.

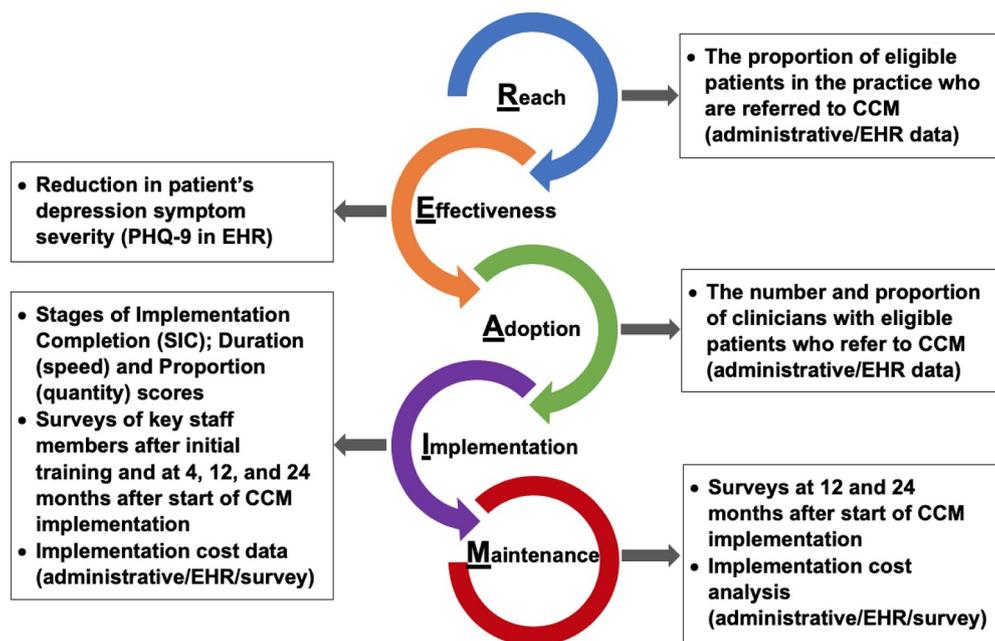


Fig. 3. Summative evaluation metrics of CCM implementation using the RE-AIM framework.

Note. CCM = Collaborative Care Model. EHR = electronic health record.

(Proportion Score). These scores can then be used in statistical analyses to understand the factors that contributed to timely stage completion, the number of stages that are important for successful program implementation by relating the SIC to other implementation outcomes, such as reach rate; and simply whether there was a degree of improvement in implementation efficiency and scale as the rollout took place. That is, were more stages completed more quickly by later sites compared to earlier ones in the rollout schedule. This analysis comprises the *implementation* domain of RE-AIM. It will be used in combination with other metrics from the EHR to determine the fidelity of implementation, which is consistent with RE-AIM.

6.3. Surveys

To understand the process and the determinants of implementation—the factors that impede or promote adoption and delivery with fidelity—a battery of surveys was administered at multiple time-points to key staff members in each practice. One challenge with large-scale implementation research is the need for measures to be both psychometrically sound as well as pragmatic. With this in mind, we adapted a set of questions for the current trial that were developed and

validated in prior studies. This low-burden assessment is comprised of items from four validated implementation surveys concerning factors at the inner setting of the organization: the Implementation Leadership Scale (Aarons et al., 2014), the Evidence-Based Practice Attitude Scale (Aarons, 2004), the Clinical Effectiveness and Evidence-Based Practice Questionnaire (Upton and Upton, 2006), and the Organizational Change Recipient's Belief Scale (Armenakis et al., 2007). In a prior study, we used confirmatory factor analysis to evaluate the four scales after shortening for pragmatism and tailoring the wording of the items (when appropriate) to the context under investigation in the study (Smith et al., under review). Further, different versions of the survey were created for administration to the various professional roles in the organization. Results showed that the scales were largely replicated after shortening and tailoring; internal consistencies were acceptable; and the factor structures were statistically invariant across professional role groups. The same process was undertaken for this study with versions of the battery developed for providers, practice leadership, support staff, and the behavioral care managers. The survey was administered immediately after initial training in the model and then again at 4, 12, and 24 months. Items were added after the baseline survey regarding the process of implementation thus far and the most

prominent barriers and facilitators to implementation of CCM in the practice. Survey-based evaluation of *maintenance* in RE-AIM, also called sustainability, will occur via the Clinical Sustainability Assessment Tool (Luke et al., 2018) to key decision makers at multiple levels in the healthcare system.

6.4. Cost of implementation

The costs incurred when adopting and delivering a new clinical intervention are a top reason attributed to lack of adoption of behavioral interventions (Glasgow and Emmons, 2007). While cost-effectiveness and cost-benefit analyses demonstrate the long-term economic benefits associated with the effects of these interventions, they rarely consider the costs to the implementer associated with these endeavors as a unique component (Ritzwoller et al., 2009). As such, decision makers value different kinds of economic evaluations, such as budget impact analysis, which assesses the expected short-term changes in expenditures for a health care organization or system in adopting a new intervention (Jordan et al., 2019), and cost-effectiveness analysis from the perspective of the implementing system and not simply the individual recipient of the evidence-based intervention being implemented (Raghavan, 2017). Eisman et al. (2019) discuss economic evaluations in implementation research.

In our study, our economic approach focuses on the cost to Northwestern Medicine to deliver CCM and will incorporate reimbursement from payors to ensure that the costs to the system are recouped in such a way that it can be sustained over time under current models of compensated care. The cost-effectiveness of CCM has been established (Jacob et al., 2012), but we will also quantify the cost of achieving salient health outcomes for the patients involved, such as cost to achieve remission as well as projected costs that would increase remission rates.

7. Conclusions

The field of implementation research has developed methods for conducting quantitative evaluation to summarize the overall, aggregate impact of implementation strategies on salient outcomes. Methods are still emerging to aid researchers in the specification and planning of evaluations for implementation studies (e.g., Smith, 2018). However, as noted in the case example, evaluations focused only on the aggregate results of a study should not be done in the absence of ongoing formative evaluations, such as in-protocol audit and feedback and other methods (see Elwy & Finley, *this issue*), and mixed and/or qualitative methods (see Hamilton and Finley, 2019). Both of which are critical for interpreting the results of evaluations that aggregate the results of a large trial and gaging the generalizability of the findings. In large part, the intent of quantitative evaluations of large trials in implementation research aligns with its clinical-level counterparts, but with the emphasis on the factors in the service delivery system associated with adoption and delivery of the clinical intervention rather than on the direct recipients of that intervention (see Fig. 1). The case example shows how both can be accomplished in an effectiveness-implementation hybrid design (see Landes et al., *this issue*). This article shows current thinking on quantitative outcome evaluation in the context of implementation research. Given the quickly-evolving nature of the field of implementation research, it is imperative for interested readers to consult the most up-to-date resources for guidance on quantitative evaluation.

Ethics approval and consent to participate

This study did not involve human subjects.

Funding

This study was supported by a grant from the Woman's Board of Northwestern Memorial Hospital and grant P30DA027828 from the National Institute on Drug Abuse, awarded to C. Hendricks Brown. The opinions expressed herein are the views of the authors and do not necessarily reflect the official policy or position of the Woman's Board, Northwestern Medicine, the National Institute on Drug Abuse, or any other part of the US Department of Health and Human Services.

CRedit authorship contribution statement

Justin D. Smith: Writing - original draft. **Mohamed Hasan:** Writing - original draft.

Declaration of Competing Interest

None.

Acknowledgments

The authors wish to thank Hendricks Brown who provided input on the development of this article and to the members of the Collaborative Behavioral Health Program research team at Northwestern: Lisa J. Rosenthal, Jeffrey Rado, Grace Garcia, Jacob Atlas, Michael Malcolm, Emily Fu, Inger Burnett-Zeigler, C. Hendricks Brown, and John Csernansky. We also wish to thank the Woman's Board of Northwestern Memorial Hospital, who generously provided a grant to support and evaluate the implementation and effectiveness of this model of care as it was introduced to the Northwestern Medicine system, and our clinical, operations, and quality partners in Northwestern Medicine's Central Region.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.psychres.2019.112521.

References

- Aarons, G.A., 2004. Mental health provider attitudes toward adoption of evidence-based practice: the Evidence-Based Practice Attitude Scale (EBPAS). *Ment. Health Serv. Res.* 6 (2), 61–74.
- Aarons, G.A., Ehrhart, M.G., Farahnak, L.R., 2014. The implementation leadership scale (ILS): development of a brief measure of unit level implementation leadership. *Implement. Sci.* 9. <https://doi.org/10.1186/1748-5908-9-45>.
- Archer, J., Bower, P., Gilbody, S., Lovell, K., Richards, D., Gask, L., ... Coventry, P., 2012. Collaborative care for depression and anxiety problems. *Cochrane Database Syst. Rev.*(10). <https://doi.org/10.1002/14651858.CD006525.pub2>.
- Armenakis, A.A., Bernerth, J.B., Pitts, J.P., Walker, H.J., 2007. Organizational change recipients beliefs scale: development of an assessment instrument. *J. Appl. Behav. Sci.* 42, 481–505. <https://doi.org/10.1177/0021886307303654>.
- Bauer, M.S., Damschroder, L., Hagedorn, H., Smith, J., Kilbourne, A.M., 2015. An introduction to implementation science for the non-specialist. *BMC Psychology* 3 (1), 32. <https://doi.org/10.1186/s40359-015-0089-9>.
- Bertrand, J.T., Holtgrave, D.R., Gregowski, A., 2009. Evaluating HIV/AIDS programs in the US and developing countries. In: Mayer, K.H., Pizer, H.F. (Eds.), *HIV Prevention*. Academic Press, San Diego, pp. 571–590.
- Brown, C.H., Curran, G., Palinkas, L.A., Aarons, G.A., Wells, K.B., Jones, L., ... Cruden, G., 2017. An overview of research and evaluation designs for dissemination and implementation. *Annu. Rev. Public Health* 38 (1). <https://doi.org/10.1146/annurev-publhealth-031816-044215>. null.
- Brown, C.H., Mohr, D.C., Gallo, C.G., Mader, C., Palinkas, L., Wingood, G., ... Poduska, J., 2013. A computational future for preventing HIV in minority communities: how advanced technology can improve implementation of effective programs. *J. Acquir. Immune Defic. Syndr.* 63. <https://doi.org/10.1097/QAI.0b013e31829372bd>.
- Brown, C.H., PoVey, C., Hjorth, A., Gallo, C.G., Wilensky, U., Villamar, J., 2015. Computational and technical approaches to improve the implementation of prevention programs. *Implement. Sci.* 10 (Suppl 1), A28. <https://doi.org/10.1186/1748-5908-10-S1-A28>.
- Cheung, K., Duan, N., 2014. Design of implementation studies for quality improvement programs: an effectiveness–cost-effectiveness framework. *Am. J. Public Health* 104 (1), e23–e30. <https://doi.org/10.2105/ajph.2013.301579>.

- Colditz, G.A., Emmons, K.M., 2017. The promise and challenges of dissemination and implementation research. In: Brownson, R.C., Colditz, G.A., Proctor, E.K. (Eds.), *Dissemination and Implementation Research in Health: Translating Science to Practice*. Oxford University Press, New York, NY.
- Curran, G.M., Bauer, M., Mittman, B., Pyne, J.M., Stetler, C., 2012. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. *Med. Care* 50 (3), 217–226. <https://doi.org/10.1097/MLR.0b013e3182408812>.
- Department of Health and Human Services, 2019. PAR-19-274: Dissemination and Implementation Research in Health (R01 Clinical Trial Optional). Retrieved from: <https://grants.nih.gov/grants/guide/pa-files/PAR-19-274.html>.
- Eisman, A.B., Kilbourne, A.M., Dopp, A.R., Saldana, L., Eisenberg, D., 2019. Economic evaluation in implementation science: making the business case for implementation strategies. *Psychiat. Res.* <https://doi.org/10.1016/j.psychres.2019.06.008>.
- Emmons, K.M., Weiner, B., Fernandez, M.E., Tu, S., 2012. Systems antecedents for dissemination and implementation: a review and analysis of measures. *Health Educ. Behav.* 39. <https://doi.org/10.1177/1090198111409748>.
- Gaglio, B., Glasgow, R.E., 2017. Evaluation approaches for dissemination and implementation research. In: Brownson, R., Colditz, G., Proctor, E. (Eds.), *Dissemination and Implementation Research in Health: Translating Science Into Practice*, second ed. Oxford University Press, New York, pp. 317–334.
- Gaglio, B., Shoup, J.A., Glasgow, R.E., 2013. The re-aim framework: a systematic review of use over time. *Am. J. Public Health* 103 (6), e38–e46. <https://doi.org/10.2105/ajph.2013.301299>.
- Glasgow, R.E., Emmons, K.M., 2007. How can we increase translation of research into practice? Types of evidence needed. *Annu. Rev. Public Health* 28, 413–433.
- Glasgow, R.E., Harden, S.M., Gaglio, B., Rabin, B., Smith, M.L., Porter, G.C., ... Estabrooks, P.A., 2019. RE-AIM planning and evaluation framework: adapting to new science and practice with a 20-year review. *Front. Public Health* 7 (64). <https://doi.org/10.3389/fpubh.2019.00064>.
- Glasgow, R.E., Klesges, L.M., Dzewaltowski, D.A., Estabrooks, P.A., Vogt, T.M., 2006. Evaluating the impact of health promotion programs: using the re-aim framework to form summary measures for decision making involving complex issues. *Health Educ. Res.* 21 (5), 688–694. <https://doi.org/10.1093/her/cyl081>.
- Glasgow, R.E., Riley, W.T., 2013. Pragmatic measures: what they are and why we need them. *Am. J. Prev. Med.* 45. <https://doi.org/10.1016/j.amepre.2013.03.010>.
- Glasgow, R.E., Vogt, T.M., Boles, S.M., 1999. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am. J. Public Health* 89 (9), 1322–1327. <https://doi.org/10.2105/AJPH.89.9.1322>.
- Grimshaw, J., Campbell, M., Eccles, M., Steen, N., 2000. Experimental and quasi-experimental designs for evaluating guideline implementation strategies. *Fam. Pract.* 17 (Suppl 1), S11–S16. https://doi.org/10.1093/fampra/17.suppl_1.s11.
- Hamilton, A.B., Finley, E.P., 2019. Qualitative methods in implementation research: an introduction. *Psychiat. Res.* 112516. <https://doi.org/10.1016/j.psychres.2019.112516>.
- Imel, Z.E., Pace, B.T., Soma, C.S., Tanana, M., Hirsch, T., Gibson, J., ... Atkins, D.C., 2019. Design feasibility of an automated, machine-learning based feedback system for motivational interviewing. *Psychotherapy* 56 (2), 318–328. <https://doi.org/10.1037/pst0000221>.
- Institute of Medicine Committee on Crossing the Quality Chasm, 2006. *Adaption to Mental Health and Addictive Disorder: Improving the Quality of Health Care for Mental and Substance Conditions*. Retrieved from Washington, D.C.
- Jacob, V., Chattopadhyay, S.K., Sipe, T.A., Thota, A.B., Byard, G.J., Chapman, D.P., 2012. Economics of collaborative care for management of depressive disorders: a community guide systematic review. *Am. J. Prev. Med.* 42 (5), 539–549. <https://doi.org/10.1016/j.amepre.2012.01.011>.
- Jordan, N., Graham, A.K., Berkel, C., Smith, J.D., 2019. Budget impact analysis of preparing to implement the family check-up 4 health in primary care to reduce pediatric obesity. *Prev. Sci.* 20 (5), 655–664. <https://doi.org/10.1007/s1121-018-0970-x>.
- Kirchner, J.E., Smith, J.L., Powell, B.J., Waltz, T.J., Proctor, E.K., 2019. Getting a clinical innovation into practice: an introduction to implementation strategies. *Psychiat. Res.* <https://doi.org/10.1016/j.psychres.2019.06.042>.
- Kronke, K., Spitzer, R., Williams, J.W., 2001. The PHQ-9. *J. Gen. Intern. Med.* 16 (9), 606–613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>.
- Landsverk, J., Brown, C.H., Smith, J.D., Chamberlain, P., Palinkas, L.A., Ogihara, M., ... Horwitz, S.M., 2017. Design and analysis in dissemination and implementation research. In: Brownson, R.C., Colditz, G.A., Proctor, E.K. (Eds.), *Dissemination and Implementation Research in Health: Translating Research to Practice*, second ed. Oxford University Press, New York, pp. 201–227.
- Lewis, C.C., Fischer, S., Weiner, B.J., Stanick, C., Kim, M., Martinez, R.G., 2015a. Outcomes for implementation science: an enhanced systematic review of instruments using evidence-based rating criteria. *Implement. Sci.* 10 (1), 155. <https://doi.org/10.1186/s13012-015-0342-x>.
- Lewis, C.C., Proctor, E.K., Brownson, R.C., 2017. Measurement issues in dissemination and implementation research. In: Brownson, R.C., Colditz, G.A., Proctor, E.K. (Eds.), *Dissemination and Implementation Research in Health: Translating Research to Practice*, second ed. Oxford University Press, New York, pp. 229–244.
- Lewis, C.C., Stanick, C.F., Martinez, R.G., Weiner, B.J., Kim, M., Barwick, M., Comtois, K.A., 2015b. The society for implementation research collaboration instrument review project: a methodology to promote rigorous evaluation. *Implement. Sci.* 10 (1), 2. <https://doi.org/10.1186/s13012-014-0193-x>.
- Luke, D.A., Calhoun, A., Robichaux, C.B., Elliott, M.B., Moreland-Russell, S., 2014. The Program Sustainability Assessment Tool: a new instrument for public health programs. *Prev. Chronic Dis.* 11, E12. <https://doi.org/10.5888/pcd11.130184>.
- Luke, D.A., Malone, S., Prewitt, K., Hackett, R., Lin, J., 2018. The Clinical Sustainability Assessment Tool (CSAT): assessing sustainability in clinical medicine settings. In: *Conference on the Science of Dissemination and Implementation in Health*. Washington, DC.
- Martinez, R.G., Lewis, C.C., Weiner, B.J., 2014. Instrumentation issues in implementation science. *Implement. Sci.* 9. <https://doi.org/10.1186/s13012-014-0118-8>.
- Powell, B.J., Stanick, C.F., Halko, H.M., Dorsey, C.N., Weiner, B.J., Barwick, M.A., ... Lewis, C.C., 2017. Toward criteria for pragmatic measurement in implementation research and practice: a stakeholder-driven approach using concept mapping. *Implement. Sci.* 12 (1), 118. <https://doi.org/10.1186/s13012-017-0649-x>.
- Powell, B.J., Waltz, T.J., Chinman, M.J., Damschroder, L.J., Smith, J.L., Matthieu, M.M., ... Kirchner, J.E., 2015. A refined compilation of implementation strategies: results from the expert recommendations for implementing change (ERIC) project. *Implement. Sci.* 10. <https://doi.org/10.1186/s13012-015-0209-1>.
- Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., ... Hensley, M., 2011. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Adm. Policy Ment. Health Ment. Health Serv. Res.* 38. <https://doi.org/10.1007/s10488-010-0319-7>.
- Rabin, B.A., Lewis, C.C., Norton, W.E., Neta, G., Chambers, D., Tobin, J.N., ... Glasgow, R.E., 2016. Measurement resources for dissemination and implementation research in health. *Implement. Sci.* 11 (1), 42. <https://doi.org/10.1186/s13012-016-0401-y>.
- Raghavan, R., 2017. The role of economic evaluation in dissemination and implementation research. In: Brownson, R.C., Colditz, G.A., Proctor, E.K. (Eds.), *Dissemination and Implementation Research in Health: Translating Science to Practice*, second ed. Oxford University Press, New York.
- Ritzwoller, D.P., Sukhanova, A., Gaglio, B., Glasgow, R.E., 2009. Costing behavioral interventions: a practical guide to enhance translation. *Ann. Behav. Med.* 37 (2), 218–227.
- Saldana, L., Schaper, H., Campbell, M., Chapman, J., 2015. Standardized measurement of implementation: the universal SIC. *Implement. Sci.* 10 (1), A73. <https://doi.org/10.1186/1748-5908-10-s1-a73>.
- Schoenwald, S., Garland, A., Chapman, J., Frazier, S., Sheidow, A., Southam-Gerow, M., 2011. Toward the effective and efficient measurement of implementation fidelity. *Adm. Policy Ment. Health Ment. Health Serv. Res.* 38. <https://doi.org/10.1007/s10488-010-0321-0>.
- Smith, J.D., 2018. An implementation research logic model: a step toward improving scientific rigor, transparency, reproducibility, and specification. *Implement. Sci.* 14 (Suppl 1), S39.
- Smith, J.D., Berkel, C., Jordan, N., Atkins, D.C., Narayanan, S.S., Gallo, C., ... Bruening, M.M., 2018. An individually tailored family-centered intervention for pediatric obesity in primary care: study protocol of a randomized type II hybrid implementation-effectiveness trial (Raising healthy children study). *Implement. Sci.* 13 (11), 1–15. <https://doi.org/10.1186/s13012-017-0697-2>.
- Smith, J.D., Li, D.H., Hirschhorn, L.R., Gallo, C., McNulty, M., Phillips, G.I., ... Benbow, N. D. (2019). Landscape of HIV implementation research funded by the national institutes of health: a mapping review of project abstracts (submitted for publication).
- Smith, J.D., Rafferty, M.R., Heinemann, A.W., Meachum, M.K., Villamar, J.A., Lieber, R. L., & Brown, C.H. (under review). Evaluation of the factor structure of implementation research measures adapted for a novel context and multiple professional roles.
- Stetler, C.B., Legro, M.W., Wallace, C.M., Bowman, C., Guihan, M., Hagedorn, H., ... Smith, J.L., 2006. The role of formative evaluation in implementation research and the Queri experience. *J. Gen. Intern. Med.* 21 (2), S1. <https://doi.org/10.1007/s11606-006-0267-9>.
- Tabak, R.G., Khoong, E.C., Chambers, D.A., Brownson, R.C., 2012. Bridging research and practice: models for dissemination and implementation research. *Am. J. Prev. Med.* 43 (3), 337–350.
- Foundation, The Health, 2015. Evaluation: What to Consider. Commonly Asked Questions About How to Approach Evaluation of Quality Improvement in Health Care. Retrieved from London, England. <https://www.health.org.uk/sites/default/files/EvaluationWhatToConsider.pdf>.
- Unützer, J., Katon, W., Callahan, C.M., Williams, J., John, W., Hunkeler, E., Harpole, L., ... IMPACT Investigators. Improving Mood-Promoting Access to Collaborative Treatment, 2002. Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. *JAMA* 288 (22), 2836–2845. <https://doi.org/10.1001/jama.288.22.2836>.
- Upton, D., Upton, P., 2006. Development of an evidence-based practice questionnaire for nurses. *J. Adv. Nurs.* 53 (4), 454–458.
- Wang, D., Ogihara, M., Gallo, C., Villamar, J.A., Smith, J.D., Vermeer, W., ... Brown, C.H., 2016. Automatic classification of communication logs into implementation stages via text analysis. *Implement. Sci.* 11 (1), 1–14. <https://doi.org/10.1186/s13012-016-0483-6>.
- Wyman, P.A., Henry, D., Knoblauch, S., Brown, C.H., 2015. Designs for testing group-based interventions with limited numbers of social units: the dynamic wait-listed and regression point displacement designs. *Prev. Sci.* 16 (7), 956–966. <https://doi.org/10.1007/s1121-014-0535-6>.
- Xiao, B., Imel, Z.E., Georgiou, P.G., Atkins, D.C., Narayanan, S.S., 2015. "Rate my therapist": automated detection of empathy in drug and alcohol counseling via speech and language processing. *PLoS One* 10 (12), e0143055. <https://doi.org/10.1371/journal.pone.0143055>.