Article

International Journal of Qualitative Methods
Volume 19: 1–11

DOI: 10.1177/1609406920916299
journals.sagepub.com/home/ijq
SAGE

Developing an Initial Program Theory to Explain How Patient-Reported Outcomes Are Used in Health Care Settings: Methodological Process and Lessons Learned

Rachel Flynn[1,2], Kara Schick-Makaroff[1], Adrienne Levay[3], and Joanne Greenhalgh[4]

Abstract

A central aspect of any theory-driven realist investigation (synthesis or evaluation) is to develop an initial program theory (IPT). An IPT can be used to frame and understand how, for whom, why, and under what contexts complex interventions work or not. Despite well-established evidence that IPTs are a central aspect to any realist investigation, there is wide variation and a lack of methodological discussion on how to develop an IPT. In this article, we present the approach that we used to develop an IPT of how patient-reported outcomes (PROs) are used in health care settings. Specifically, we completed a systematic review to extract tacit theories reported in the literature. The benefit of this approach was that it provided a rigorous review of the literature in the development of IPTs. The challenges included (1) rediscovering what is already well established in the theoretical literature, (2) generating an overabundance of partial candidate theories, and (3) extensive use of time and resources for what was the first stage to our larger funded research study. Our recommendations to other scholars considering this approach are to ensure that they (1) live within their means and (2) narrow the scope of the research question and/or develop a conceptual framework using middle-range theories. These methodological insights are highly relevant to researchers embarking on a realist investigation, tasked with developing an IPT.

Keywords

realist synthesis, initial program theory, program theories, patient-reported outcomes, patient-reported outcome measures, patient-reported experience measures, kidney disease

Background

Realist syntheses are driven by the question: "What works, how, for whom, in what circumstances and to what extent?" (Pawson, 2002; Pawson et al., 2005). It is a theory-driven approach to evidence synthesis philosophically rooted in scientific realism (Pawson, 2006). A central aspect of any realist synthesis is to develop, test, and refine a program theory. A "program theory" articulates the ideas and assumptions underlying how and why an intervention or program is expected to work (Pawson, 2006). A program theory can identify and map out the key components of the program, the outcomes the program is intended to generate, and the contexts that might shape the mechanisms through which the program works that contribute to particular outcomes (Muckumbang et al., 2017; Pawson, 2002; Pawson et al., 2005).

Developing an initial program theory (IPT) is the first step in a theory-driven realist research cycle (Muckumbang et al., 2017). An IPT "sketches the terrain" that will be investigated and assists in refining the scope for realist inquiry (synthesis or evaluation; Wong et al., 2013). Through realist inquiry, an IPT

[1] Faculty of Nursing, University of Alberta, Edmonton, Canada
[2] Faculty of Nursing, Edmonton Clinic Health Academy, University of Alberta, Edmonton, Canada
[3] Alberta Data Institute, Alberta Innovates, Edmonton, Canada
[4] Sociology and Social Policy, University of Leeds, United Kingdom

Corresponding Author:
Rachel Flynn, Faculty of Nursing, Edmonton Clinic Health Academy, University of Alberta, 11405-87 Ave., Edmonton, Alberta, Canada T6G 1C9.
Email: rachel.flynn@ualberta.ca

is tested and refined to more accurately explain how and why interventions work on the ground. While adoption of realist synthesis appears to be an increasingly popular approach for evidence syntheses of complex interventions, there is a paucity of literature that discusses or provides clear examples of how to undertake this necessary first step in a theory-driven realist research cycle. We attend to this gap in the literature and propose that more detailed methodological guidance will support rigorous development of IPTs with clear application of realist principles. In this article, we first introduce existing methodological guidance and approaches for IPT development. Next, we discuss the approach that we employed to develop an IPT for our realist synthesis on how patient-reported outcomes and experience measures (jointly referred to here as PROs) can be used to enhance person-centered nephrology care (Schick-Makaroff et al., 2017, 2019). Finally, we conclude with a discussion of the benefits, challenges, and lessons learned from the approach we employed, offering recommendations to realist researchers planning and undertaking IPT development.

Aim

The purpose of this article is to describe the benefits and challenges of the approach we used to develop an IPT for a realist synthesis to explain how patient-reported outcomes (PROs) are used in health care settings. We offer transferable lessons to advance existing knowledge and provide a methodological example of IPT development for researchers undertaking a realist synthesis.

Realist Syntheses

Realist syntheses aim to identify the underlying causal generative mechanisms that explain "how" outcomes are caused under certain contexts that produce outcomes of a program. Realist syntheses are guided by the context (C) + mechanism (M) = outcome (O) configuration (CMOc; Pawson & Tilley, 1997). A CMOc is a hypothesis that the program works (O) because of the action of some underlying mechanism (M), which only comes into operation in certain contexts (C; Pawson & Manzano-Santaella, 2012, p. 184). At an individual level, programs or interventions offer resources to participants and how these resources change the reasoning or responses of participants constitute the mechanisms that explain how things work beneath what the eye can see (Dalkin et al., 2015; Pawson, 2006). Westhorp (2018) notes that mechanisms can operate at different levels of a system (not just the individual) but fundamentally are causal forces or processes that explain how an intervention does or does not work.

Existing Methodological Guidance for IPT Development

Realist synthesis is still a relatively new approach that is a part of growing methodological advancement in evidence reviews. In recent years, the RAMESES project provided training materials and guidance on realist synthesis and developing IPTs (Wong et al., 2013). Section 4 of the RAMESES training materials focuses on (1) developing and refining realist program theory, (2) quality standards for program theories, (3) examples from the literature, and (4) a learning activity (Wong et al., 2013). The development of IPTs for realist investigation occurs through a variety of ways such as conducting workshops with stakeholders, reviewing program documentation (policy documents, funding applications, program descriptions, and so on) or by reviewing a small selection of literature about the program type (Wong et al., 2013). The variation as to how an IPT can be developed presents challenges for students and researchers alike embarking on a realist synthesis or evaluation. The RAMESES training materials for realist synthesis outline a list of six questions to assist with selecting an appropriate approach for developing an IPT and constructing and refining it (Table 1).

Table 1. Questions to Assist Constructing and Refining a Realist Program Theory.

Question 1	Do you need to construct a realist program theory for each outcome of interest? If not, why not?
Question 2	What sources and resources (e.g., other researchers, experts, service users) will you draw on to help you develop your realist program theory?
Question 3	What processes will you develop and put in place to develop and iteratively refine your program theory/theories?
Question 4	Are there existing substantive theories that are will help to inform your program theory/theories?
Question 5	What assumptions are built into the program theory? What assumptions are we (the reviewers) making? What assumptions are there in the data? Which ones do we need to challenge and why?
Question 6	What data, from where, might help to test and refine the theory?

Source: Wong et al. (2013, p. 28).

The RAMESES (2014) training materials also provide quality standards for realist synthesis that range from inadequate to excellent for constructing and refining a realist program theory. An excellent standard is defined to show that:

> The relationship between the program theory and relevant substantive theory is identified. Implications of the final theory for practice, and for refinements to substantive theory where appropriate, are described. The final realist program theory comprises multiple context-mechanism-outcome configurations (describing the ways different mechanisms fire in different contexts to generate different outcomes) and an explanation of the pattern of CMOs. (RAMESES, 2014, p. 4)

This quality standard is helpful for researchers to ensure that their IPT is explicit and follows realist principles. Despite this guidance, there is variation in the literature on how people develop an IPT for realist inquiry. For example, some researchers develop IPTs through realist synthesis (Groot et al., 2017), qualitative research (Mukumbang et al., 2017), other forms of

knowledge synthesis (Lhussier et al., 2016), or by using/further developing an explicit program theory that already existed (MacDonald et al., 2016). Shearn et al. (2017) provide four potential approaches for developing IPTs: (1) using concepts from abstract theories which are used to inform current or comparable interventions, (2) using concepts from abstract theory which are selected purposively for the research synthesis or evaluation by the research team but which have not been referenced in the program literature, (3) extracting tacit theories about what is working and why from interventions on similar topics, reported in the literature, and (4) extracting tacit theories (about what is working and why) directly from stakeholders via one-to-one interviews, brainstorming, documentation of the current intervention, and/or developed by the research team who may use their own experiential or professional knowledge.

The approach we used aligns with Shearn et al.'s (2017) third suggestion, where we conducted a systematic literature search and extracted tacit theories about "what is working and why" from similar interventions. Using the realist CMOc heuristic, we aimed to identify how contexts (i.e., health care settings) shape mechanisms (the processes, reasoning, or behaviors triggered by the use of PROs) through which the intervention (PRO feedback) brings about an outcome (at individual and aggregate levels). Shearn et al. (2017) allude to some potential challenges to this chosen approach of using data-driven approaches only, such as (a) identifying what is already well established in the existing literature, (b) generating an overabundance of candidate theories, and (c) developing theory that may be unstructured. In this article, we add further specificity and discuss some of these challenges that we experienced. To reflect on and provide insight into the process of our IPT development, we draw upon the methods and approach that we employed, meeting notes from our weekly core research team meetings during IPT development, our team report, and our own sense-making processes.

Research Context

Developing IPT for PROs in Kidney Care

The objectives of our study were to (1) understand theories that explain how PROs are used and (2) develop a kidney-specific program theory about use of PROs in nephrology that may enhance person-centered care by testing and refining the theory through a realist synthesis of the empirical literature (Schick-Makaroff et al., 2019). Our intention was to first informally search the literature to identify existing theories on how PROs are used in healthcare settings to develop our IPT. To locate these theories, we (a) undertook a systematic review of the peer-reviewed literature to identify existing theories and (b) consulted our methodological expert (J.G.), research team practitioners (specifically nephrologists), and our patient advisory committee (PAC) to solicit their feedback for IPT refinement. Research team members identified three exemplar papers a priori including Greenhalgh (2009), Santana and Feeny's framework (2014), and Valderas and Alonso (2008).

Method

IPT Development Through Systematic Review

Following the seven stages for realist synthesis (Wong et al., 2013), we set out to create IPTs to be tested and refined through evidence in order to meet objective one of our study: To formulate IPTs that explain how patient-reported outcomes (PROs) are used in health care settings (Schick-Makaroff et al., 2017, 2019). Patient-reported outcome and experience measures (PROMs and PREMS) are regarded globally as a means for people to report on the impact of illness on their health and quality of life as well as their experience with care (Greenhalgh, 2009). More specifically, PROMs are self-report instruments used to obtain self-appraisals about outcomes relevant to a person's quality of life (e.g., well-being; overall health; symptoms; functional status; and other aspects of psychological, social, and spiritual quality of life; Fayers & Machin, 2016). PREMs refer to "questionnaires measuring the patients' perceptions of their experience whilst receiving care" (Kingsley & Patel, 2017, p. 137). Although PREMs and PROMs are different types of instruments, both comprise important outcomes of health care, we refer to them jointly here as patient-reported outcomes (PROs; Schick-Makaroff et al., 2019). Development of our IPTs closely aligns with modern perspectives of PRO measurement validity which increasingly emphasizes the importance of focusing on "use" in contributing toward measurement validity evidence (Hawkins et al., 2018). Thus, IPT development on use of PROs contributes to the field of measurement validation of PROs.

The IPT development process took 15 months (December 2016–March 2018) with three paid part-time research assistants. To begin, and with the aid of a library scientist, a systematic strategy (Additional File 1) was created to search the peer-reviewed literature for two concepts: PRO measures and theory. Including primary research, theoretical or review literature, we retrieved 13,412 articles. After removing duplicated, 6,295 records were title screened, and 1,210 abstracts were screened by the core research team using NVivo Version 11 and EndNote X7 (with 10% double check). Inclusion/exclusion criteria were developed iteratively, trialed and discussed in weekly core research team meetings, and they are reported elsewhere (Schick-Makaroff et al., 2019). Full text of 42 articles were screened by three reviewers. Likewise, 10% of the included full texts were double-checked. Thirty-four full-text articles were included for extraction for our IPTs (see Figure 1).

Prior to our extraction process, we drew from six exemplary articles on PRO use in health care settings (Greenhalgh, 2009; Greenhalgh et al., 2014; Porter et al., 2016; Santana & Feeny, 2014; Valderas & Alonso, 2008; Velikova et al., 2004) and worked through inductively extracting CMOc. This created an opportunity for us as a team to understand and define what we meant by context, mechanism, and outcome in relation to

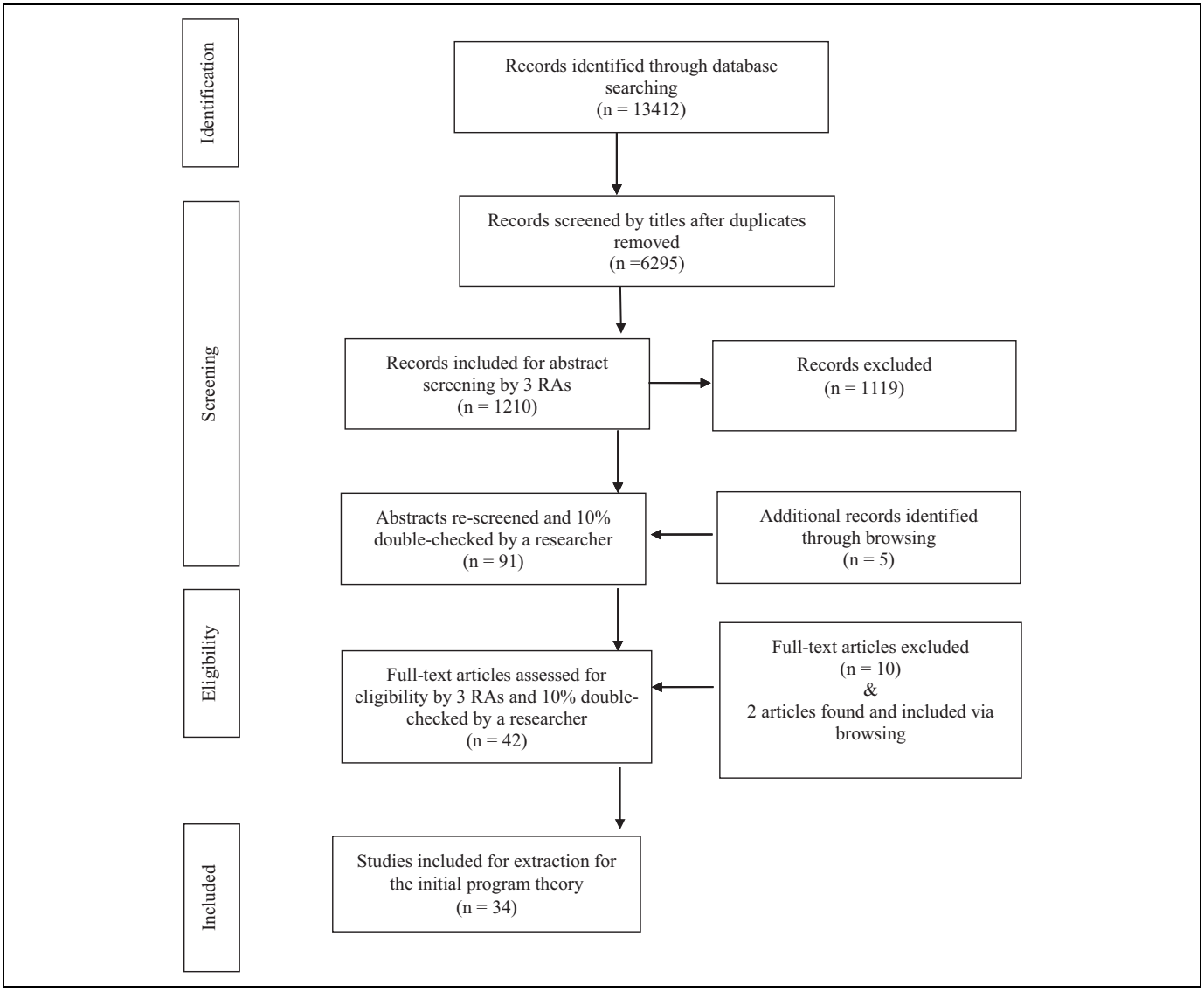

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses. Screening process for the initial program theory (Schick-Makaroff et al., 2019).

how PRO feedback works. For the IPT data extraction process, we created a table in Microsoft word and began abductive processes by creating "if-then" statements from which we developed complete CMOc. We trialed this process with the core research team, discussed and revised on a weekly basis. This process occurred over 3 months.

The core research team then divided into two groups: individual and aggregate levels of decision making. We defined individual levels of decision making as PRO use at point of care by patients, caregivers, and health care providers. We defined aggregate levels of decision making, as PRO use at government, policy, and system-wide levels. In each group, CMOc were inductively extracted from the 34 articles at individual and aggregate levels by a primary reviewer and then the secondary reviewers, and input into tabular format using Microsoft word. A tertiary reviewer reviewed one third of the articles. (The pattern for the selection of the articles, which were numbered 1–34, was the sequence of multiples of 3 [e.g., 14, 17, 20, ...]). All extractions were reviewed either by the individual or aggregate team or at least one member of the other team. The core question that guided this process was: "Does this provide theoretical explanation about how PROs are used at an individual or aggregate level in health settings?"

Simultaneously, as we gathered partial CMOc inductively from our included studies, we also abducted from theory and ideas to form more complete CMOc which we later compared, tested, and refined through another realist synthesis in the nephrology literature.

The extraction process was extensive. We met weekly to update on progress and discuss questions. Combining tables from the teams was also a large undertaking that resulted in a 170-page Microsoft word document of individual and

aggregate CMOc or partial CMOc. To ensure consistency between and among teams for data extraction, we developed a data extraction codebook/dictionary and we created "nickname themes" for our CMOc while under development so that team members could quickly reference them in discussion. We also created a process for extracting quotes: if direct quote, use quotation marks, page number, and reference. If not direct quote, page number and reference. Our final IPT CMOc are provided in Tables 2 and 3.

Team Meetings and Other Work

In addition to article screening and extraction of CMOc or partial CMOc of the 34 included articles, our core research team met weekly. Our weekly meetings involved development of our protocol and PROSPERO registration, our first database searches, realist training workshops with our methodological expert (J.G.), IPT work, and refined program theory work conducted through a review of existing literature. Contextually, at the time of our IPT development, one of the research team members (J.G.) was also undertaking a realist synthesis on PRO use in health care. A protocol was published at the beginning of our systematic search (Greenhalgh et al. 2014), and the full report was published at the end or our extraction (Greenhalgh, Dalkin, Gibbons, et al., 2017). While our work was at different stages, there was productive overlap between our program theories.

Patient Advisory Committee

We engaged with a PAC during our IPT development process. We consulted PAC members who included people living with kidney disease (on home dialysis, in-hospital dialysis, kidney transplant recipients, and predialysis) and their spouses. One patient was a part of the full research team from the beginning of the research proposal development. In May 2016, the PAC members and research team participated in a 1-day long Canadian Institute for Health Research (CIHR) Strategy for Patient-Orientated Research (SPOR & CIHR, n.d.) Training that was tailored to our team's work together on the project. We met again in August 2018 to provide an update and to solicit their input on our very early IPT development. We also reported our IPTs to the PAC in March 2018 and discussed program theories in October 2018.

Findings

The search and screening results of our systematic review for IPT development are presented through an adapted Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram (Figure 1). Our IPT approach resulted in six CMOc at an individual level (Table 2) and eight CMOc at an aggregate level (Table 3). Our IPTs map out the main components of using PROs in health care settings, anticipated outcomes, contexts that might shape mechanisms that produce these outcomes, and sequencing of these elements at individual and aggregate levels.

To further test and refine our IPTs, we subsequently undertook a realist synthesis of the kidney-specific peer-reviewed and gray literature to address the second objective of our study. Our kidney-specific program theory was later developed and tested as an evidence-based theory about use of PROs in nephrology for enhancing person-centered care at individual and aggregate levels of decision making. The final program theory was dramatically different from the IPTs presented in this article but could not have been developed without this foundational work.

Challenges to IPT Development Approach

Reflecting on the approach we undertook to develop our IPT, below we have summarized three key challenges that we encountered: (1) rediscovering what is already well-established in the theoretical literature, (2) generating an overabundance of partial candidate theories, and (3) extensive use of time and resources for what was the first stage to our larger funded research study. Our IPT development work took 15 months and required the resources of a research librarian, three paid research staff, realist training, and weekly core research team meetings.

First, from focusing on peer-reviewed literature alone, we were faced with the tension of rediscovering what was potentially already known in the field or duplicating other work underway. Contextually, we were apparent of this tension because of other realist syntheses projects on similar topics which were being undertaken separately and in different stages (Greenhalgh, Dalkin, Gibbons, et al., 2017; Greenhalgh, Dalkin, Gooding, et al., 2017). Nevertheless, our larger research team decided that to ensure our IPTs were rigorous and fully informed future program theory development using kidney resources, we needed to undertake a systematic investigation of the literature. While the process offered reassurance and alignment with the field, it was not without significant use of resources.

Second, due to the magnitude of our search records (n = 6,295) and the final number of included documents (n = 34), with two teams conducting inductive data extraction of contexts, mechanisms, and outcomes, we ended up with a 170-page document. Extractions from articles were rarely complete CMOc, resulting in "theory fragments" or partial CMOc. We faced the challenge of abductively synthesizing these into CMOc, so that we could move forward with our realist review to test and refine our IPTs. We started broadly with an overabundance of partial CMOc and then faced the challenge of abductive refinement, deciding which were the most relevant, substantiated, and rigorous to test and refine. We chose to take 14 IPTs forward to refinement and testing in the next phase of our study.

Third, as mentioned, we had a large database of records to screen, and our screening process contained several steps. The title and abstract screening involved six stages, including

Table 2. Initial Program Theory: Individual Level Use of PROs in Health Care.

Nickname	CMO Configuration	Supporting References
"Valuable-by-association"	#1. If PRO measures are available, in conjunction with other clinical measures (context), then clinicians may believe they are clinically important, sufficient, and usable (i.e., comprehensive; reasoning) to detect symptoms/conditions, monitor current treatment, influence clinical decision making, and lead to changes in patient care/management (outcomes). However, clinicians may limit PRO use (outcome) if they mistrust PRO data (context and/or reasoning) or feel it is not comprehensive enough to inform clinical decision making (context and/or reasoning). [a]	Ayers et al. (2013), Bartlett and Ahmed (2017), Bartlett et al. (2017), Bingham et al. (2017), Boyce et al. (2014), Bruckel et al. (2015), Chen et al. (2013), Desantis et al. (2016), Feldman-Stewart & Brundage (2009), Greenhalgh et al. (2005), Greenhalgh, Dalkin, Gibbons, et al. (2017), Greenhalgh, Dalkin, Gooding, et al. (2017), Hunter et al. 2015, Porter et al. (2016), Santana and Feeny (2014), Santana et al. (2015)
"Ease of collection and use—1"	#2. If a PRO is completed and fed back (i.e., relevant, useful, and meaningful) to clinicians (resource) alongside adequate support and resources (resource or preexisting context), then this may trigger patients' and clinicians' desires to engage/discuss/address issues (reasoning), which in turn, may facilitate/improve patient–provider communication/shared decision-making/satisfaction/relationship and outcomes (outcomes). However, there may be a difference in priorities/decisions between patients and providers regarding which issues to discuss (context), leading to changes in patient management and/or emotional distress experienced by patients (outcome).	Ayers et al. (2013), Bartlett & Ahmed (2017), Bingham et al. (2017), Catania et al. (2016), Chen et al. (2013), Donaldson, (2008), Feldman-Stewart and Brundage (2009), Greenhalgh et al. (2005), Greenhalgh (2009), Greenhalgh et al. (2014), Greenhalgh, Dalkin, Gibbons et al. (2017), Greenhalgh, Dalkin, Gooding, et al. (2017), Hunter et al. (2015), Jayadevappa and Chhatre (2011), Noonan et al. (2017), Porter et al. (2016), Richardson et al. (2015), Santana and Feeny (2014), Valderas and Alonso (2008), Velikova et al. (2004)
"Electronic collection"	#3. If electronic collection of PRO data is well designed (i.e., clinically feasible and relevant, fits into clinical workflow, user-friendly; resource) along with organizational support and resources (resources and/or preexisting contexts), then this may facilitate PRO use (outcome), enhance patient self-management (outcome), and support research activities (outcome) due to electronically enhanced "usability," "interpretability," and "actionability" (reasoning).	Ahmed et al (2017), Ayers et al. (2013), Bartlett and Ahmed (2017), Bingham et al. (2017), Bruckel et al. (2015), Desantis et al. (2016), Donaldson (2008); Greenhalgh, Dalkin, Gooding, et al. (2017), Locklear et al. (2017), Nelson et al. (2015), Noonan et al. (2017), Porter et al. (2016), Santana et al. (2015)
"Ease of collection and use—2"	#4. If the application of PROs fits into the existing clinical workflows and feedback is clinically relevant, timely, and provided over multiple times (resource), then this may trigger providers' beliefs [b] and use of PRO data (reasoning) to inform clinical decision-making (outcome). However, concerns about practical constraints may affect providers' responses.	Ayers et al. (2013), Bartlett and Ahmed (2017), Bingham et al. (2017), Boyce et al. (2014), Bruckel et al. (2015), Catania et al. (2016), Chen et al. (2013), Desantis et al. (2016), Donaldson (2008), Greenhalgh et al. (2005), Greenhalgh (2009), Greenhalgh et al. (2014), Greenhalgh, Dalkin, Gooding, et al. (2017), Hunter et al. (2015), Nelson et al. (2015), Noonan et al. (2017), Porter et al. (2016), Santana and Feeny (2014), Santana et al. (2015), Schünemann et al. (2006)
"Information and education—providers"	#5. If providers receive education on how to interpret PROs (resource), then this will change their use of PRO information (outcome). However, education alone will not change behavior. Training, feedback, and administrative support (resource) may facilitate PRO use by providers with enhanced engagement and ownership (reasoning) as a routine part of care, and enable change in behavior (outcome). This may lead to supported patient engagement (outcome), enhanced shared-decision making between the provider and patient (outcome), and PRO use as part of routine care (outcome).	Boyce et al. (2014), Chen et al., (2013), Greenhalgh et al., (2005), Greenhalgh, Dalkin, Gooding, et al. (2017), Noonan et al. (2017), Porter et al. (2016), Santana and Feeny (2014), Santana et al. (2015)
"Information and education—patients"	#6. If PRO results are routinely provided to patients (or identified by patients) for self-management (resource), with accompanying education (resource), then not only will patient self-management improve along with their well-being (outcome), but PROs implementation will be more successful (outcome).	Catania et al. (2016), Chen et al. (2013), Greenhalgh, Dalkin, Gooding, et al. (2017), Hunter et al. (2015), Noonan et al. (2017), Santana et al. (2015)

Note. Contexts (C), different types of mechanisms (M; i.e., resources, reasoning, etc.), and outcomes (O) are identified in brackets within CMO configurations. In initial program theory, all CMO components are not required to be present. PRO = patient-reported outcome.

[a] Clinicians' perspectives of the credibility of PROs may influence their use.

[b] Providers' beliefs to be further developed in program theory.

Table 3. Initial Program Theory: Aggregate Level Use of PROs in Health Care.

Nickname	CMO Configuration	Supporting References
"Patient choice"	#1. If patients use publicized PROs as performance data to compare and choose their service providers and opt to receive care at higher performing hospitals (resource), then this will put pressure on lower performing hospitals (reasoning) to improve their care (outcome).	Greenhalgh (2009), Greenhalgh et al. (2014), Greenhalgh, Dalkin, Gibbons, et al. (2017)
"Reputation"	#2. If the feedback (i.e., through media) about performance data is released publicly (resource), then this may trigger concerns and response by providers about their reputation (response). However, if there is no change in market share (outcome/context), then providers may not respond (response) because patients make their decisions based on factors other than performance data (context). Therefore, there may not be a change in patients' choice of service providers (outcome).	Greenhalgh et al. (2014), Greenhalgh, Dalkin, Gibbons, et al. (2017), Greenhalgh, Dalkin, Gooding, et al. (2017)
"Competition"	#3. If providers use feedback of publicized performance data (i.e., that are clinically led; resource) to reflect on their practice (response), compare (response), and compete with their peers (response), then this may trigger providers' motivations (response) to improve performance (outcome), thus enhancing providers' reputation (outcome), and increasing patient referrals (outcome).	Boyce et al. (2014), Greenhalgh, Dalkin, Gibbons, et al. (2017), Greenhalgh, Dalkin, Gooding, et al. (2017)
"Carrots"	#4. If providers are rewarded based on the feedback of publicized performance data (resource), then this may encourage a sense of shared responsibility (reasoning). However, the perceived creditability of data (context), concerns about loss of incentives, and reimbursement or accreditation (response) may lead to unintended consequences from financial incentives or detraction from providing quality of care (outcome).	Boyce et al. (2014), Greenhalgh, Dalkin, Gooding, et al. (2017), Hunter et al. (2015), Jayadevappa and Chhatre (2011), Noonan et al. (2017)
"Valuing data"	#5. If feedback of publicized performance data is perceived to be credible, appropriate, and applicable to practice (context), then this may trigger providers' trust, value (response), and use of performance data (outcome).	Boyce et al. (2014), Greenhalgh (2009), Greenhalgh, Dalkin, Gibbons, et al. (2017), Greenhalgh, Dalkin, Gooding, et al. (2017), Locklear et al. (2017), Porter et al. (2016), Greenhalgh, Dalkin, Gibbons, et al. (2017), Greenhalgh, Dalkin, Gooding, et al. (2017), Reeve et al. (2013)
"Organizational support"	#6. If the feedback of performance data is valued by providers (context) alongside of adequate resources and support (context), then this may enable providers' ability to act on data (response) for quality improvement (outcome) when system- and organizational-wide strategies are in place (resource/context). However, concerns about practical issues (context) may influence providers' responses (response).	Bartlett and Ahmed (2017), Boyce et al. (2014), Chen et al. (2013), Desantis et al. (2016), Donaldson (2008), Greenhalgh, Dalkin, Gibbons, et al. (2017), Greenhalgh, Dalkin, Gooding, et al. (2017), Hunter et al. (2015), Locklear et al. (2017); Mamiya et al. (2017); Porter et al. (2016); Reeve et al. (2013)
"Learning health system"	#7. If PRO data are routinely collected and generated for different purposes (resource), then this may contribute to the creation of a learning health system (outcome).	Ahmed et al. (2017), Bartlett et al. (2017), Bingham et al. (2017), Chen et al. (2013), Greenhalgh et al. (2014), Greenhalgh, Dalkin, Gibbons, et al. (2017), Santana and Feeny (2014), Valderas and Alonso (2008)
"Electronic collection and sharing"	#8. If PRO data are collected and shared electronically (i.e., databases or personal health records; resource), then care may be more timely, equitable, improved (outcome), and in turn support population health/regulatory agencies working toward efficient allocation of resources (outcome).	Bartlett et al. (2017), Greenhalgh (2009), Mamiya et al. (2017)

Note. Contexts (C), different types of mechanisms (M; i.e., resources, reasoning, etc.), and outcomes (O) are identified in brackets within CMO configurations. PRO = patient-reported outcome.

practice screening, title screening, abstract screening using NVivo for a computer-assisted approach, and abstract screening via EndNote X7, 10% double-checking, and rescreening the included abstracts. We also completed "practice" screening and met as a core research team to discuss our questions to aid team members new to realist synthesis. While our approach

enabled a highly rigorous systematic process, it is important to note the challenges. Our systematic approach to IPT development took longer than we anticipated, resulting in the use of more resources, including time and costs, than expected. Thus, our full project resulted in two systematic reviews—one to develop our IPTs and another to test and further refine our IPTs.

Discussion

Recommendations for IPT Development

Reflecting on the processes we undertook to develop our IPT, we have gained lessons and insights which we offer in the following two recommendations.

Recommendation 1: "Live within your means."

Anecdotally and from our team's firsthand experience, it must be acknowledged that embarking on a full systematic search to develop realist IPT can be resource-intensive. To counter this, Saul et al. (2013) proposed a framework for conducting Rapid Realist Reviews. Even for an undertaking like this that nominally promises to mitigate the issue of resource constraints, they still require significant resources. For example, they suggest an ideal rapid realist review team would include a project manager, a local reference group, a librarian, a review team consisting of two to four individuals, a synthesis lead, and an academic or research lead. Not only is payment to all these individuals something to consider, whether through salaries, in-kind, or service agreements, but so too is the amount of time that each of these types of team members may be able to reasonably provide. The key is to balance available labor (i.e., graduate students who can only work part-time vs. full-time nonstudent staff), available funds, expertise, and time as well as to discuss expectations of the comprehensiveness of the IPT. It is also possible to develop strategies to help optimize the resources available for a realist synthesis. For example, we found that having a well-defined intervention and inclusion/exclusion criteria (Schick-Makaroff et al., 2017, 2019) enabled us to have clear guidance for data screening which increased the efficiency at which we were able to screen articles as well as help standardize the selection process across the different reviewers on the team. Full-team discussions about how to "live within your means" may help mitigate both anticipated an unanticipated challenge in the development of IPTs.

Recommendation 2: "Narrow the scope of the research question and/or develop a conceptual framework using middle-range theories (MRTs)."

Having an overabundance of partial IPTs challenged our team to explain patterns or fully unpack mechanisms. This overabundance may have been due to (a) not having created sufficiently narrow boundaries around our research question or (b) not integrating MRTs until refinement of our IPTs.

First, focusing reviews is given an entire section in the realist training materials created by Wong et al. (2013). Even for one specific intervention (in our case, use of PROs in health care), the questions that could be asked about it are infinite. From a realist perspective, this is a characteristic of complexity. Wong et al. (2013) suggest that containing a review is therefore important and can be done in a number of ways including narrowing the question, narrowing the aspect of the intervention to be investigated, narrowing the scale (e.g., by countries or cultures or timeframes), and mitigating expectations about the comprehensiveness of the review.

Second, Shearn et al. (2017) incorporate the "use of concepts from abstract theories" into two of their suggestions for how to approach developing IPTs. For realist researchers, these concepts from abstract theories can be MRTs. MRTs can be used to develop an a priori set of contexts, mechanisms, and outcomes with which to reduce the variety of CMOs identified from the data in the literature, from interview transcripts, and/or document analysis. For our team, the integration of MRTs earlier on could have provided a framework to understand how, why, and in what circumstances the PROs are used in health care settings, thus creating parameters around what "nuggets" of data from the literature should be extracted and included in IPT development. One mitigating strategy we used to manage our overabundance of partial CMOc and to help with sense-making was to create two succinct categories, individual- and aggregate-level PRO use. Research team discussions to critically appraise the scope of the research question, as well as integration of MRTs may facilitate IPT development in realist investigation.

Limitations

A limitation that we encountered was that only two of the six members had engaged in this type of realist CMO extraction before, and only three of the six had ever participated in realist work—so developing our IPT was also our training to realist synthesis and extraction. Using the RAMSES training materials for realist synthesis, we only drew upon Questions 1–3 and 6 in-depth when first developing our IPT. We did not consider Question 5: "What assumptions are built into the program theory? What assumptions are we (the reviewers) making? What assumptions are there in the data? Which ones do we need to challenge and why?" nor Question 6: "What data, from where, might help to test and refine the theory?" until we were refining our IPTs. This may be considered a limitation to our IPT development.

Conclusion

In this article, we present the approach that we used to develop an IPT of how PROs are used in health care settings. We believe that reporting on the process we took to develop our IPT can improve transparency and future guidance on IPT development. Developing IPTs through systematic processes are both rigorous and detailed; however, this approach should

be embarked on with caution, resources, and the needed skill set. We recommend the use of existing program theories (where available) to improve coherence and data abstraction. Further, earlier integration of MRTs can mitigate the challenges faced when developing IPT from solely data-driven approaches. These methodological insights are highly relevant to graduate students and researchers embarking on a realist investigation, tasked with developing IPTs in future realist work.

Authors' Note

Research Team members involved in this research include Stephanie Thompson, Richard Sawatzky, Onouma Thummapol, Scott Klarenbach, and Mehri Karimi-Dehkord. Our patient advisory committee members involved in this research include Loretta Lee (cochair), Tammy Fifield, Justin LeRoux, Jon Gilchrist, Sandra Johnstone, Marcus Alphonsus Walsh, Julia Walsh, Jeff Costley, Suzanna Pomrenke, and Beulah Fitzpatrick.

Acknowledgments

Rachel Flynn is supported by a Women and Children's Health Research Institute postdoctoral fellowship. Kara Schick-Makaroff is a new investigator funded by Can-SOLVE CKD—KRESCENT (Canadians Seeking Solutions and Innovations to Overcome Chronic Kidney Disease; Kidney Research Scientist Core Education and National Training Program).

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Study funded by Kidney Foundation of Canada, Allied Health Research Grant, KFOC 160024.

ORCID iD

Rachel Flynn https://orcid.org/0000-0003-2980-7441

Supplemental Material

The supplemental material for this article is available online.

References

Ahmed, S., Ware, P., Gardner, W., Witter, J., Bingham, C., Bartlett, S., & Kairy, D. (2017). Montreal accord on patient-reported outcomes (PROs) use series—Paper 8: Patient-reported outcomes in electronic health records can inform clinical and policy decisions. *Journal of Clinical Epidemiology*, *89*, 160–167.

Ayers, D. C., Zheng, H., & Franklin, P. D. (2013). Integrating patient-reported outcomes into orthopaedic clinical practice: Proof of concept from force-TJR. *Clinical Orthopaedics and Related Research*, *471*(11), 3419–3425.

Bartlett, S., & Ahmed, S. (2017). Montreal accord on patient-reported outcomes (PROs) use series—Paper 1: Introduction. *Journal of Clinical Epidemiology*, *89*, 114–118.

Bartlett, S., Ahmed, S., Witter, J., & Cella, D. (2017). Montreal accord on patient-reported outcomes (PROs) use series—Paper 6: Creating national initiatives to support development and use—The PROMIS example. *Journal of Clinical Epidemiology*, *89*, 148–153.

Bingham, C., Bartlett, S., Noonan, V., Auger, C., Feldman, D., & Ahmed, S. (2017). Montreal accord on patient-reported outcomes (PROs) use series—Paper 4: Patient-reported outcomes can inform clinical decision making in chronic care. *Journal of Clinical Epidemiology*, *89*, 136–141.

Boyce, M. B., Browne, J. P., & Greenhalgh, J. (2014). Surgeon's experiences of receiving peer benchmarked feedback using patient-reported outcome measures: A qualitative study. *Implementation Science*, *9*, 84.

Bruckel, J., Wagle, N., O'Brien, C., Elias, J., McKenna, S., Meyers, P., Fifer, M., Pomerantsev, E., & Yeh, R. W. (2015). Feasibility of a tablet computer system to collect patient-reported symptom severity in patients undergoing diagnostic coronary angiography. *Critical Pathways in Cardiology: A Journal of Evidence-Based Medicine*, *14*(4), 139–145.

Canadian Institutes for Health Research & Strategy for Patient-Oriented Research. (n.d.) http://www.cihr-irsc.gc.ca/e/41204.html

Catania, G., Beccaro, M., Costantini, M., Timmins, F., Zanini, M., Aleo, G., Bagnasco, A., & Sasso, L. (2016). What are the components of interventions focused on quality-of-life assessment in palliative care practice? A systematic review. *Journal of Hospice and Palliative Nursing*, *18*(4), 310–316.

Chen, J., Ou, L., & Hollis, S. J. (2013). A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. *BMC Health Services Research*, *13*, 211.

Dalkin, S. M., Greenhalgh, J., Jones, D., Cunningham, B., & Lhussier, M. (2015). What's in a mechanism? Development of a key concept in realist evaluation. *Implement Science*, *10*(49), 1–7.

Desantis, D., Baverstock, R. J., Civitarese, A., Crump, R. T., & Carlson, K. V. (2016). A clinical perspective on electronically collecting patient-reported outcomes at the point-of-care for overactive bladder. *Canadian Urological Association Journal*, *10*(11–12), E359–E366.

Donaldson, M. S. (2008). Taking PROs and patient-centered care seriously: Incremental and disruptive ideas for incorporating PROs in oncology practice. *Quality of Life Research*, *17*(10), 1323–1330.

Fayers, P. M., & Machin, D. (2016). *Quality of life: The assessment, analysis and reporting of patient-reported outcomes*. Wiley-Blackwell.

Feldman-Stewart, D., & Brundage, M. D. (2009). A conceptual framework for patient-provider communication: A tool in the PRO research tool box. *Quality of Life Research*, *18*(1), 109–114.

Greenhalgh, J. (2009). The applications of pros in clinical practice: What are they, do they work, and why? *Quality of Life Research*, *18*(1), 115–123.

Greenhalgh, J., Dalkin, S., Gibbons, E., Wright, J., Valderas, J. M., Meads, D., & Black, N. (2017). How do aggregated patient-reported outcome measures data stimulate health care improvement? A realist synthesis. *Journal of Health Services Research & Policy*, *23*(1), 57—65.

Greenhalgh, J., Dalkin, S., Gooding, K., Gibbons, E., Wright, J., Meads, D., Black, N., Valderas, J. M., & Pawson, R. (2017). Functionality and feedback: A realist synthesis of the collation, interpretation and utilisation of patient-reported outcome measures data to improve patient care. *Health Services and Delivery Research*, *5*(2), 1–280.

Greenhalgh, J., Long, A. F., & Flynn, R. (2005). The use of patient reported outcome measures in routine clinical practice: Lack of impact or lack of theory? *Social Science & Medicine*, *60*(4), 833–843.

Greenhalgh, J., Pawson, R., Wright, J., Black, N., Valderas, J. M., Meads, D., Gibbons, E., Wood, L., Wood, C., Mills, C., & Dalkin, S. (2014). Functionality and feedback: A protocol for a realist synthesis of the collation, interpretation and utilisation of proms data to improve patient care. *BMJ Open*, *4*(7), e005601.

Groot, G., Waldron, T., Carr, T., McMullen, L., Bandura, L. A., Neufeld, S. M., & Duncan, V. (2017). Development of a program theory for shared decision-making: A realist review protocol. *Systematic Reviews*, *6*. https://doi.org/10.1186/s13643-017-0508-5

Hawkins, M., Elsworth, G. R., & Osborne, R. H. (2018). Application of validity theory and methodology to patient-reported outcome measures (PROMS): Building an argument for validity. *Quality of Life Research*, *27*(7), 1695–1710. https://doi-org.login.ezproxy.library.ualberta.ca/10.1007/s11136-018-1815-6

Hunter, C., Fitzpatrick, R., Jenkinson, C., Darlington, A. S., Coulter, A., Forder, J. E., & Peters, M. (2015). Perspectives from health, social care and policy stakeholders on the value of a single self-report outcome measure across long-term conditions: A qualitative study. *BMJ Open*, *5*(5), e006986.

Jayadevappa, R., & Chhatre, S. (2011). Patient centered care—A conceptual model and review of the state of the art. *Open Health Services and Policy Journal*, *4*, 15–25.

Kingsley, C., & Patel, S. (2017). Patient-reported outcome measures and patient-reported experience measures. *BJA Education*, *17*(4), 137–144.

Lhussier, M., Carr, S. M., & Forster, N. (2016). A realist synthesis of the evidence on outreach programs for health improvement of traveller communities. *Journal of Public Health*, *38*(2), e125–e132.

Locklear, T., DeBar, L. L., Willig, J., Rundell, S., Blackhall, L., Zatzick, D., Staman, K., Bhavsar, N., Weinfurt, K., & Abernethy, A. P. (2017). Case studies from the clinic: Initiating and implementing patient-reported outcome measures. *eGEMs*, *5*(1):7.

MacDonald, M., Pauly, B., Wong, G., Schick-Makaroff, K., van Roode, T., Wilson Strosher, H., Kothari, A., Valaitis, R., Manson, H., O'Briain, W., Carroll, S., Lee, V., Tong, S., Dickenson-Smith, K., & Ward, M. (2016). Supporting successful implementation of public health interventions: Protocol for a realist synthesis. *BMC Systematic Reviews*, *5*. https://doi.org/10.1186/s13643-016-0229-1

Mamiya, H., Ahmed, S., Buckeridge, D., Lix, L., Gardner, W., & Bartlett, S. (2017). Montreal accord on patient-reported outcomes (pros) use series—Paper 5: Patient-reported outcomes can be linked to epidemiologic measures to monitor populations and inform public health decisions. *Journal of Clinical Epidemiology*, *89*, 142–147.

Mukumbang, F. C., Van Belle, S., Marchal, B., & Van Wyk, B. (2017). Exploring 'generative mechanisms' of the antiretroviral adherence club intervention using the realist approach: A scoping review of research-based antiretroviral treatment adherence theories. *BMC Public Health*, *17*(1), 385.

Nelson, E. C., Eftimovska, E., Lind, C., Hager, A., Wasson, J. H., & Lindblad, S. (2015). Patient reported outcome measures in practice. *BMJ (Online)*, *350*, g7818.

Noonan, V. K., Lyddiatt, A., Ware, P., Jaglal, S. B., Riopelle, R. J., Bingham, C. O., Figueiredo, S., Sawatzky, R., Santana, M., & Bartlett, S. J. (2017). Montreal accord on patient-reported outcomes (PROS) use series—Paper 3: Patient-reported outcomes can facilitate shared decision-making and guide self-management. *Journal of Clinical Epidemiology*, *89*, 125–135.

Pawson, R. (2002). Evidence-based policy: In search of a method. *Evaluation*, *8*(2), 157.

Pawson, R. (2006). *Evidence-based policy: A realist perspective*. Sage.

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2005). Realist review—A new method of systematic review designed for complex policy interventions. *Journal of Health Services Research & Policy*, *10*, 21–34.

Pawson, R., & Manzano-Santaella, A. (2012). A realist diagnostic workshop. *Evaluation*, *18*(2), 176–191.

Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. Sage.

Porter, I., Goncalves-Bradley, D., Ricci-Cabello, I., Gibbons, C., Gangannagaripalli, J., Fitzpatrick, R., Black, N., Greenhalgh, J., & Valderas, J. M. (2016). Framework and guidance for implementing patient-reported outcomes in clinical practice: Evidence, challenges and opportunities. *Journal of Comparative Effectiveness Research*, *5*(5), 507–519.

RAMESES. (2014). Quality standards for realistic synthesis, 1–8. http://www.ramesesproject.org/media/RS_qual_standards_researchers.pdf

Reeve, B. B., Wyrwich, K. W., Wu, A. W., Velikova, G., Terwee, C. B., Snyder, C. F., Schwartz, C., Revicki, D. A., Moinpour, C. M., Lyons, J. C., Lenderking, W. R., Hinds, P. S., Hays, R. D., Greenhalgh, J., Gershon, R., Fenny, D., Fayers, P. M., Cella, D., Brundage, M., & Butt, Z. (2013). ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Quality of Life Research*, *22*(8), 1889–1905.

Richardson, M. M., Grobert, M. E., & Meyer, K. B. (2015). Randomized controlled trials 3: Measurement and analysis of patient-reported outcomes. *Methods in Molecular Biology*, *1281*, 191–206.

Santana, M. J., & Feeny, D. (2014). Framework to assess the effects of using patient-reported outcome measures in chronic care management. *Quality of Life Research*, *23*(5), 1505–1513.

Santana, M. J., Haverman, L., Absolom, K., Takeuchi, E., Feeny, D., Grootenhuis, M., & Velikova, G. (2015). Training clinicians in how to use patient-reported outcome measures in routine clinical practice. *Quality of Life Research*, *24*(7), 1707–1718.

Saul, J. E., Willis, C. D., Bitz, J., & Best, A. (2013). A time-responsive tool for informing policy making: Rapid realist review. *Implementation Science*, *8*(103), 1–15.

Schick-Makaroff, K., Sawatzky, R., Thompson, S., Klarenbach, S., & Flynn, R. (2017). Patient-reported outcomes in clinical kidney

practice (PRO kidney): A realist synthesis. PROSPERO: International prospective register of systematic reviews. Registration crd42017056063. http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42017056063

Schick-Makaroff, K., Thummapol, O., Thompson, S., Flynn, R., Karimi-Dehkordi, M., Klarenbach, S., Sawatzky, R., & Greenhalgh, J. (2019). Strategies for incorporating patient-reported outcomes in the care of people with chronic kidney disease (PRO kidney): Protocol for a realist synthesis. *BMC Systematic Review*, *8*(20), 1—10.

Schünemann, H. J., Akl, E. A., & Guyatt, G. H. (2006). Interpreting the results of patient reported outcome measures in clinical trials: The clinician's perspective. *Health & Quality of Life Outcomes*, *4*, 62.

Shearn, K., Allmark, P., Piercy, H., & Hirst, J. (2017). Building realist program theory for large complex and messy interventions. *International Journal of Qualitative Methods*, *16*, 1–11.

Valderas, J. M., & Alonso, J. (2008). Patient reported outcome measures: A model-based classification system for research and clinical practice. *Quality of Life Research*, *17*(9), 1125–1135.

Velikova, G., Booth, L., Smith, A. B., Brown, P. M., Lynch, P., Brown, J. M., & Selby, P. J. (2004). Measuring quality of life in routine oncology practice improves communication and patient well-being: A randomized controlled trial. *Journal of Clinical Oncology*, *22*(4), 714–724.

Westhorp, G. (2018). Understanding mechanisms in realist evaluation and research. In N. Emmel (Ed.), *Doing realist research* (pp 41–57). Sage.

Wong, G., Westhorp, G., Pawson, R., & Greenhalgh, T. (2013). *Realist synthesis: RAMESES training materials*. http://www.ramesesproject.org/index.php?pr=Project_outputs