MAIN ARTICLE

Rigorously interpreted quotation analysis for evaluating causal loop diagrams in late-stage conceptualization

Andrada Tomoaia-Cotisel,[a,b]* Samuel D. Allen,[c] Hyunjung Kim,[d] David Andersen[e] and Zaid Chalabi[f,g]

Abstract

As a field, system dynamics has made more progress in developing formal model evaluation methods that use quantitative data than ones that use qualitative data. This article describes a formal method for evaluating a causal loop diagram (CLD) in late-stage conceptualization – referred to as rigorously interpreted quotation analysis – and illustrates its application in a case study. The method uses a systematic and explicit interpretive process to confirm or disconfirm all diagram elements in a CLD by comparing it to stakeholders' verbatim descriptions of their experiences in a complex dynamic situation. In so doing, this method enables the resolution of discrepancies between a CLD and qualitative data, building confidence in the structural aspects of a dynamic hypothesis. It does so via a process that is approachable for experts and stakeholders alike.

Syst. Dyn. Rev. (2022)

Introduction

In the allegory of the blind people and the elephant, blindness *hinders* people from accurately perceiving the elephant they encounter. Each person touches only a part of the body, such as tusk, tail, or side, and believes that an elephant is like a sword, snake, or wall. After much disagreement, they conclude: all parts must be considered together to learn the truth.

Similarly, in social problems, our cognitive limitations *hinder* us from accurately understanding a system's structure and resulting dynamic behavior. In system dynamics, we hope that we can create a more *holistic*

[a] RAND Corporation, Santa Monica, CA, USA
[b] Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, London, UK
[c] Department of Operations and Industrial Engineering, The School of Business, Worcester Polytechnic Institute, Worcester, MA, USA
[d] Thomas Family Fellow in Business, Department of Management, College of Business, California State University, Chico, CA, USA
[e] Rockefeller College of Public Affairs and Policy, University at Albany—State University of New York, Albany, NY, USA
[f] Institute for Environmental Design and Engineering, The Bartlett Faculty of the Built Environment, University College London, London, UK
[g] Department of Public Health, Environments and Society, Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, London, UK

* Correspondence to: Andrada Tomoaia-Cotisel, RAND Corporation, 1776 Main St, Santa Monica, CA 90401, USA. E-mail: andrada@rand.org

Accepted by Luis Felipe Luna-Reyes, Received 8 June 2021; Revised 10 November 2021 and 22 January 2022; Accepted 25 January 2022

System Dynamics Review
System Dynamics Review
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/sdr.1701

understanding by (i) eliciting individuals' experiences and (ii) diagramming and simulating what has been shared. Validation techniques enable us to *interrogate* the elicited data as we evaluate and revise this understanding be it in the form of a diagram or a simulation model.

Just like our problem stakeholders, we are one of the blind people, with our own perception of the elephant – a perception that is limited and inaccurate. Through the modeling process, we seek to improve our collective perception by bringing together the available, relevant data and correcting that perception as we go.

As a field, system dynamics has made more progress in developing formal model-evaluation methods that use quantitative data than ones that use qualitative data (Andersen *et al.*, 2012). Formal evaluation methods that make use of high-quality qualitative data are needed. Innovating formal tools that employ qualitative data will improve the field's ability to triangulate and check our work against empirical data to systematically mitigate the risk of bias introduced in the modeling process (Sterman, 2018c). By creating explicit links between model elements and qualitative data and by systematically documenting the interpretive process, system dynamics modelers can build confidence in models while uncovering flaws in modeler interpretation that can lead to further structural modifications. Formal methods used by qualitative researchers outside system dynamics offer useful insights, but they are not designed with precise causal diagramming conventions in mind. Therefore, the method introduced in this article attempts to fill the need (inside system dynamics) for formal evaluation methods that use high-quality qualitative data and the need (outside system dynamics) for such methods to use precise causal diagramming conventions.

The rigorously interpreted quotation (RIQ) method is an analytical tool that supports systematic analysis of qualitative data. In this article, we present examples for conducting RIQ analysis in the conceptualization phase of a system dynamics modeling process (although the method can also be used for other purposes (c.f., (Tomoaia-Cotisel, 2018)). As will be further explained later in the article, RIQ analysis builds on existing qualitative methods in system dynamics and in the broader social-science literature.

What is the rigorously interpreted quotation method?

The RIQ method is one way of implementing *structure-validating processes* (Barlas, 1996; Lane, 1995; Richardson, 2019; Saeed, 1992) where comparison and reconciliation of system structures takes place. More specifically, the RIQ method uses: (i) a causal loop diagram (CLD) developed during initial conceptualization and (ii) purposive text[i] regarding the problem being

[i]According to Kim and Andersen (2012), purposive text data comes from stakeholders with extensive knowledge about the system or problem at hand. Its descriptions are "frank and unfeigned," and they reveal mental models of the stakeholders "with a reasonable degree of confidence" (pp. 312–313).

DOI: 10.1002/sdr

studied. Purposive text is used to evaluate[ii] and modify the CLD during late-stage conceptualization. The product of RIQ analysis is an improved version of the project's conceptualized system. This CLD becomes the basis of the working model later used for *behavior validating processes* (Lane, 1995; Richardson, 2019; Saeed, 1992). See Figure A1 in the online supporting information for how this method fits into a system dynamics process. (See the "future research" section for additional proposed applications of the RIQ method.)

Motivation for development of the rigorously interpreted quotation method

The RIQ method was initially developed in Tomoaia-Cotisel (2018) for a case study of primary-care transformation involving the use of CLDs and system dynamics simulation. This study attempted to understand why it has been so challenging for many primary-care clinicians inside one academic medical center to transform their clinics into ones that deliver comprehensive, coordinated, continuous, and accessible care for their full panel of patients. Using a system dynamics model, the author showed how the health-care system structure, stakeholder preferences, and policies *together* shape the variety of trajectories experienced.

In this case study, which we refer to as *the parent study*, stakeholder preferences and their mental models of the system played a key role in determining the system's behavior. The author needed formal methods to elicit individual mental models, combine these understandings to create a CLD, and evaluate (and make any needed revisions to) that CLD *before* it was used as the blueprint for simulation modeling and policy analysis. The RIQ method was developed for this verification purpose.

Organization of the article

Our article is organized as follows: First, we review relevant literature in system dynamics and in social science more generally – literature on different qualitative methods that influenced the development of RIQ analysis. We also discuss the gap that the RIQ method attempts to fill. Second, we briefly introduce the research context in which the RIQ method was originally developed and used. The goal is to illustrate where in the research process the RIQ method was applied and how it contributed to the overall research outcomes. Third, using simple examples from that research project, we detail how RIQ analysis can be conducted. Fourth, we provide a summary of CLD refinement results from RIQ analysis to illustrate potential value of using the RIQ method. Fifth, we conclude by reflecting on contributions and

[ii]When thinking of *evaluating* a CLD, people might use different terms (e.g., to *verify* a CLD or to *validate* a CLD).

DOI: 10.1002/sdr

limitations of the current approach and propose future research areas. In so doing, we present RIQ analysis as a useful guide for a broad range of researchers and practitioners working with qualitative data and CLDs.

Literature review

In this section, we briefly review the system dynamics literature and the broader literature on qualitative methods in social sciences which influenced the RIQ method development. But before delving into this literature, we discuss the importance of qualitative data in modeling and how it relates to the reliability and validity issues in qualitative research.

Qualitative model validation and research quality

Written and mental databases play an important role in system dynamics research (Forrester, 1961, 1980). Since *Industrial Dynamics* (Forrester, 1961), "descriptive information" has been essential for all stages of modeling—encompassing model conceptualization, formulation, experimentation, and evaluation. Forrester provided two reasons supporting the use of both quantitative and qualitative data in the treatment of validity: (i) many aspects of system dynamics model development are based on qualitative *mental data*, and the same type of data should be used to validate such a model, and (ii) it expands the number of problems which can be studied. In Forrester's words: as "a preponderant amount of human knowledge is in nonquantitative form… [therefore] model building and model validation do not [need to] stop at the boundary where numerical data fail" (p. 129).

Similar calls for using nonnumerical data for validation can be found in the broader field of modeling and simulation. For example, operations researchers Oral and Kettani (1993) reaffirm Forrester's observation that the richest source of data for modeling is "mental data bases" consisting of mental models, and that such data have a place in model validation (p. 226). Of the formal methods for qualitative validation, these authors describe system dynamics' inherent advantages in conceptualization and conclude that, when it is "using the cognitive capacities of the relevant actors fully [it] is perhaps the most promising [method] … for determining the validity of a given 'conceptual model'" (p. 227).

Nonnumerical data are made up of various types of qualitative information that are found jumbled in terms of their forms, topics, sources, styles, and purposes. Qualitative analysis involves sorting out the data that are more informative for the study purpose from the rest (Forrester, 1980; Guest *et al*., 2012a; Hodgkinson and Clarkson, 2005; Kim and Andersen, 2012; Van Maanen, 1979). Qualitative researchers across the social sciences have developed a variety of formal interpretive processes in order to analyze qualitative

DOI: 10.1002/sdr

data, and they can be applied with varying degrees of quality (Harding and Seefeldt, 2013).

According to Creswell (2014),[iii] quality in qualitative research combines both validation and reliability. Validation is about using multiple activities that enhance the researcher's ability to check the accuracy of findings (cf. Chapter 9). As the purpose of validation in qualitative research is *to convince* researchers, participants, and/or readers of their findings' validity, it is a social, judgmental process; this is in line with the system dynamics notion of validity as a qualitative, social, confidence-building process (Lane, 2015). Reliability is about having a consistent and transparent process. A consistent process would include double checking the transcription of recorded discussions as well as sharing and evaluating coding (Creswell, 2014). A transparent process involves reporting the process's step-by-step procedures, so that others can follow (Yin, 2009).

As we hope to illustrate in this article, the RIQ method contributes both to the validity and reliability of the modeling process. It provides a structured approach to assessing the accuracy of findings and enhancing buy-in from problem owners (clients, stakeholders) and interested readers. Furthermore, the RIQ method provides a formal procedure for researchers to use reliably across projects and research team members. In the following section, we discuss how the RIQ method draws from and improves upon the existing qualitative techniques used in system dynamics.

Relating rigorously interpreted quotation analysis to four formal qualitative techniques in system dynamics validation

In system dynamics, the development of more formal processes for using qualitative data to validate the structure of conceptual and simulation models has long been perceived to be an important, yet challenging task (Barlas, 1996; Luna-Reyes and Andersen, 2003). Before relating RIQ analysis to processes and standards in social science more generally, we will first explore how it relates to four formal qualitative techniques used in system dynamics: (i) the inductive system diagrams method, (ii) purposive text analysis, (iii) the disconfirmatory interview, and (iv) group model building.

Inductive system diagrams

The inductive system diagrams method (Burchill and Kim, 1993; Burchill and Fine, 1994; Burchill and Fine, 1997) develops a CLD using grounded theory-based coding of field notes (see also "strategy 1" in deGooyert's review of SD methods in organizational theory (de Gooyert, 2018)). After grounded theory coding is completed, one CLD is developed for each key

[iii]Other authors suggest a more nuanced concept of trustworthiness when discussing validity (Creswell and Miller, 2000) and reliability in qualitative research (Crabtree and Miller, 1999).

DOI: 10.1002/sdr

variable with its associated links. Before combining these CLDs together, the diagram is evaluated against qualitative data to ensure that all links and loops are supported by evidence, that the diagram explains the study situation, and that the variables share a common level of abstraction (Burchill and Kim, 1993). The RIQ method shares the similar model-validation activity of ensuring each link and loop is supported by evidence. Unlike the inductive system diagrams method, RIQ analysis is designed to occur after one CLD has already been developed and uses raw transcript data rather than coded field notes. Also, the inductive system diagrams method mentions comparing the diagram with data in a "review" (Burchill and Kim, 1993, p. 14), and authors using this method describe this step as "returning to the data" (Perlow *et al.*, 2002, p. 934). This process is described as involving a "constant comparison" where revisions to a diagram result when a data element that is compared with the diagram points to a revision such that the resulting diagram contains the "accumulated knowledge" and "clearly and concisely display[s] the current state of accumulated evidence and inferences" (Burchill and Fine, 1997, pp. 469–470). The RIQ method can support and improve this comparison process by providing a *transparent* and *systematic* tool, allowing analysts to document and present what they have found and allowing stakeholders to evaluate evidence supporting diagram elements as well as how analysts interpreted that evidence.

Purposive text analysis

Purposive text analysis (Eker and Zimmermann, 2016; Kim and Andersen, 2012; Turner *et al.*, 2013) uses qualitative data in its text form to develop stock and flow diagrams. Purposive text analysis assumes availability of reliable data (i.e. purposive text) that can provide rich information about the system structure, and model elements emerge from the data through a systematic process of coding. By documenting clear linkages between model elements and their source texts, purposive text analysis allows the coder's interpretive process to be traced and recorded. While this may present an opportunity for models to be validated against their source data (Turner *et al.*, 2013), this method has been used primarily in early conceptualization of model development (cf. Eker and Zimmermann, 2016; Turner *et al.*, 2013). The RIQ method expands this application to the later stage of model conceptualization by coding a fresh set of data against the existing conceptual model, looking for information that either validates or disconfirms model elements.

The disconfirmatory interview

The disconfirmatory interview (Andersen *et al.*, 2012) is a formal method for working with stakeholders to validate simulation-model structure and behavior. Interviews with individual stakeholders involve the presentation

DOI: 10.1002/sdr

of key model features showing structure and behavior alongside textual descriptions, or narratives, describing structural assumptions and the elicitation of participants' input for what needs to be changed in the model. Andersen *et al.* (2012) explain common challenges faced by disconfirmatory interviews as (i) the need to validate system dynamics models with nonmodelers, (ii) ensuring correct interpretation of stakeholder input, and (iii) the need to have live interviews. While the RIQ method shares similar model validation goals with disconfirmatory interview techniques, it may offer ways to overcome the described challenges. For example, the RIQ method utilizes descriptions of stakeholders' experiences as shared in their natural language and therefore does not require the stakeholders to understand the modeling language. Also, the RIQ method involves a thorough documentation of the modeler's interpretative process, allowing its assessment to take place if needed. Furthermore, the RIQ method does not require the interviewee to be present while the validation takes place and therefore offers much flexibility in the research process. Finally, the RIQ method evaluates models at the conceptualization stage, something which has yet to be explored in the disconfirmatory interview literature. It is also worth noting that the RIQ method focuses both on *confirmation and* disconfirmation whereas the Disconformity Interview focuses strongly of the latter (Andersen *et al.*, 2012).

Group model building

Group model building (Andersen *et al.*, 1997; Hovmand, 2014; Richardson and Andersen, 1995; Vennix, 1999) engages stakeholder groups throughout the modeling process, and this practice allows model formulation to take place implicitly with model validation. Group model building scripts (Andersen and Richardson, 1997; Hovmand *et al.*, 2012) illustrate how to select key stakeholders and how to design sessions to elicit valuable data for building and assessing a model. The use of documented group model building processes allows best practices to be implemented in a consistent and systematic way. In group model building, the role of an experienced facilitator is critical in reducing bias resulting from counterproductive group/power dynamics and maintaining a healthy level of disagreement in open conversations (Andersen *et al.*, 2012; Vriens and Achterbergh, 2006). While group model building is a preferred modeling method for many modelers, holding multiple group model building sessions throughout a project may not be feasible due to various organizational constraints. In such cases, the RIQ method can complement the modeling process by facilitating model formulation and validation where session transcripts, interview transcripts, and other qualitative data provided by stakeholders are available. Furthermore, the RIQ method can support *explicit* evaluation in group model building,

DOI: 10.1002/sdr

which has been a challenge due to difficulty of documenting and curating the evidence behind structures generated during sessions (Farr, 2017).

Summary

In sum, the system dynamics field has developed several formal processes that leverage the strengths of qualitative methods for model evaluation. With the growing use of CLDs,[iv] the need is as great as ever to address the call made by Richardson:

> [The] future could prompt us to develop a more robust catalogue of qualitative 'methods'–ways of using qualitative maps in complex dynamic situations that minimise the potential for spurious insights and maximise the likelihood that practitioners will know what can be reliably inferred from a given map without further analysis, and when further work, presumably involving modelling and computer simulation, is required. (Richardson, 1999, p. 441)

> … the increasing use of qualitative maps will pressure us toward creating a set of more or less reliable *principles of qualitative system dynamics…* (Richardson, 1999, p. 442, emphasis added)

Urging further progress on this front, Sterman (2018c) called for the field to continue "innovat[ing] to develop new methods that are appropriate for the models we build" (p. 36) p.36) and that promote "rigor, reliability, relevance and impact" (pp. 7, 37). With respect to the uses of qualitative data and testing conceptual models, he calls for the use of rigorous methods, grounded in formal social-science techniques to mitigate the risk of bias entering from qualitative data[v] or from the modeler[vi] (pp. 26–27) and for qualitative diagrams to be based in evidence and rigorously tested (p. 39). This echoes previous calls for cross-pollination with other social-science methods (Kopainsky and Luna-Reyes, 2008; Luna-Reyes and Andersen, 2003).

We believe the RIQ method we introduce in this study builds toward these ambitious goals. Based on the above literature, we see a need for formal tools

[iv] A search for "causal loop diagram" on App.Dimentions.AI shows a steady increase in publications using this term since 2011.

[v] Mitigating the risk of bias from qualitative data does not mean that qualitative data is inherently more prone to bias than quantitative data. For example, on the quantitative side, bias can be introduced by what the analyst decides to measure and how they decide to measure it. To mitigate bias from qualitative data, it is helpful to consider data segmentation in validation efforts (see the "sampling considerations" section below). It is also helpful for the reader to have the opportunity to see examples of the qualitative data on which the model is based, to see for themselves whether it is indeed purposive text. And, while not the subject of this article, Tomoaia-Cotisel (2018) shows how RIQ analysis can also be used to assess saturation – the point at which we have gathered enough qualitative data to be reasonably confident that we have heard enough of what we needed to hear.

[vi] Mitigating the risk of bias from the modeler does not mean removing the modeler from the process, rather it means transparently documenting the modeler's perspective so that the modeler and others can evaluate (at least some of) the qualitative data and modeler's interpretation and thus have the opportunity to disconfirm that interpretation.

DOI: 10.1002/sdr

to confirm and disconfirm model elements by interrogating reliable purposive text data. Such tools would be capable of dissecting stakeholder experiences as they are shared in their natural language. Such tools would also facilitate documenting and curating the evidence to transparently and systematically ensure that each link and loop is supported by evidence.

Relating rigorously interpreted quotation analysis to processes and standards in social science more generally

As indicated above, formalizing tools for qualitative research involves going beyond system dynamics literature. Below, we briefly discuss standards and methods used by other social-science approaches using similar data for similar goals; specifically, we explore: (i) mixed methods for causal inference, (ii) thematic analysis, (iii) causal mapping, and (iv) sampling considerations.

Mixed methods for causal inference

Mixed-methods (Miller *et al.*, 2013) studies are often commissioned to build understanding of both the *structure* and the *dynamics* of social systems and to do so in an *empirically rigorous* way. For example, Miller *et al.* (2013) argue that their purpose is to "*fully* capture the *complex* interactions among components, including interactions among multiple levels of analysis and *over time*" (p. 2125, emphasis added). Mixed-methods studies integrate qualitative and quantitative elements in study design, methods, and interpretation (Fetters *et al.*, 2013; Miller *et al.*, 2013; Tashakkori and Teddlie, 1998). When used for causal inference, mixed methods can be seen as triangulation that includes not just multiple methods but also multiple sources of evidence (Reiss, 2009). Qualitative evidence for complex causal processes and unmeasured factors provides a useful complement to quantitative analysis concerned with causality (Harding and Seefeldt, 2013). This evidence often comes in qualitative form. Best practice involves the use of *process tracing* (developing causal chains from detailed descriptions) and *pattern matching* (evaluating the extent to which behavior descriptions match those predicted by a given causal link) (Andersen). By using system dynamics diagramming conventions, RIQ analysis facilitates precisely tracing complex dynamic processes and interactions between variables as expressed in verbal descriptions of experiences.

Thematic analysis

Thematic analysis uses *chunks* of text data as a proxy for stakeholders' experience with the problem being studied (Guest *et al.*, 2012b; Ryan and Bernard, 2000). The concepts and relationships envisioned may include causal links between variables, but more often, broader types of relationships are considered (Ryan and Bernard, 2000). One *fundamental* task shared by

DOI: 10.1002/sdr

all formal thematic analysis is "constructing models (relationships among codes) and testing these models against empirical data" (Ryan and Bernard, 2000, p. 274). Validation involves looking for so-called *negative cases* that can "disconfirm parts of a model or suggest new connections" (p. 278). An early application of this *negative case analysis* method was in the classic work of Becker *et al.* (1961) which suggests that negative case analysis involves two stages: in the first stage, researchers continually recraft a "provisional hypothesis" (p. 28). In the second stage, they evaluate a "final statement" (p. 39) of the hypothesis using the "careful inspection of all negative instances" (p. 39). These negative instances include ones where part of the hypothesis is omitted in a quotation and the researchers believe that this is due to lack of knowledge, and others where elements different from the hypothesis are identified. Each of these latter cases is evaluated to see if it represents an outlier or if it merits revising the hypothesis (p. 44). However, if the first stage is done right, there are usually few negative cases remaining unaccounted for (p. 39). RIQ analysis as used in this article is akin to evaluating the final statement and facilitates careful inspection by using a systematic tool designed for system dynamics diagramming conventions. RIQ analysis encompasses negative case analysis by identifying instances of disconfirmation and omission (whether by ignorance, willful omission, or simplified description). RIQ analysis also complements negative case analysis by documenting evidence that confirms diagram structure.

Causal mapping

Causal mapping (Nakayama and Armstrong, 2005) is an interdisciplinary field of social scientists interested in mapping participants' descriptions of causality in the form of diagrams. In this field, validating causal maps derived from primary interview transcripts is referred to as a "vexed question" (Hodgkinson and Clarkson, 2005, p. 53) due to common data-quality problems—especially for unstructured text data—such as "sentence fragments, incomplete thoughts, and over-elaborate explanations" (p. 53). Instead of using such data, researchers prefer to directly elicit participants' input for what needs to be changed in the model by showing them maps (Armstrong, 2005). RIQ analysis is designed with the understanding that people's experiences with complex dynamic systems do not naturally produce tidy descriptions of causal structure. RIQ analysis confronts this problem directly and addresses it successfully. For example, Table 4 below provides an example of how it does so for an overelaborate quotation. The RIQ method prepares text data before causal analysis, sorts out specific phrases that relate to specific variables in a quotation, and shows the modeler's interpretation for the causal structure identified. RIQ analysis also provides more flexibility by relying on descriptions of stakeholders' experiences

DOI: 10.1002/sdr

as shared in their natural language rather than having to ask stakeholders to engage with maps.

Sampling considerations

Before we close, we provide a note on sampling[vii] as it relates to qualitative methods for diagram evaluation. Sampling is a relevant consideration when assessing the internal and external validity of findings (in this case, in the form of a diagram/simulation and derived insights). In ethnographic decision modeling for example, interview data are used to develop decision trees(Ryan and Bernard, 2006). The internal validity of the model is assessed using a second set of qualitative data from the same population that was used in developing the model(Ryan and Bernard, 2006). The external validity of the model is assessed using a third set of qualitative data from other populations (Ryan and Bernard, 2006). To our knowledge, the (system dynamics and) social science methods described above have largely been used to improve and/or assess internal validity of diagrams and/or simulation models. RIQ analysis is able to assess both internal and external validity (depending on the sampling design of the study) so long as the data that is collected is purposive text.

Summary

In sum, these social-science approaches have developed formal processes that permit comparing qualitative data to hypothesized mechanisms. However, similar to the qualitative methods in system dynamics, these methods present gaps that RIQ analysis can fill. We need a tool that facilitates careful *and systematic* inspection for confirmation as well as disconfirmation, relying on descriptions of stakeholders' experiences as shared in their natural language. RIQ analysis does so using precise causal diagramming conventions.

Research context for method development

Figure 1 below shows how the RIQ method was used in relation to other formal qualitative and quantitative techniques in the parent study. The parent study was undertaken as a dissertation (Tomoaia-Cotisel, 2018). (See Figure A1 in the online supporting information for how this method fits into a system dynamics process.)

[vii]Some readers may be more familiar with sampling in quantitative analysis, where quantitative data is selected from a larger pool of data. Similar to quantitative analysis, sampling in qualitative analysis refers to how study sites and participants are chosen from the larger pool of sites and participants experiencing the problem being investigated.

DOI: 10.1002/sdr

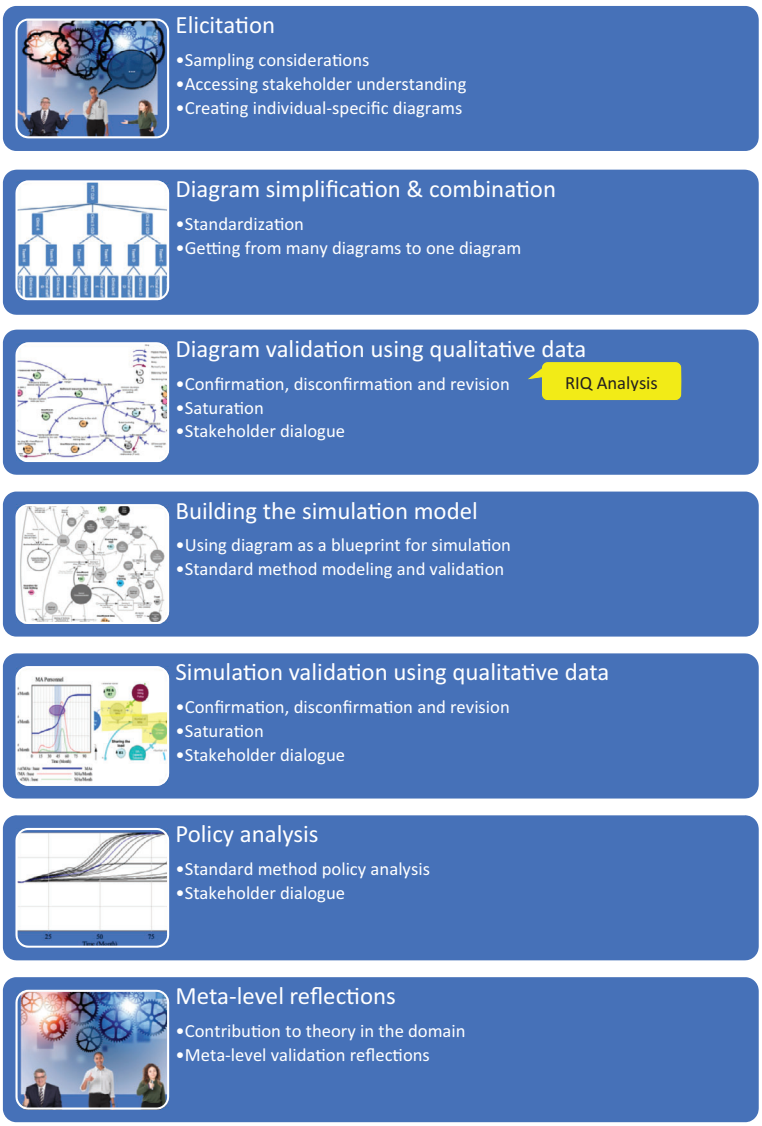

Fig. 1. Location of RIQ Analysis in the Parent Study. This figure shows the location of RIQ analysis in the parent study using a yellow call-out box next to "confirmation, disconfirmation and revision." The methods used (and the results obtained) in the parent study are described further in Tomoaia-Cotisel (2018) [Color figure can be viewed at wileyonlinelibrary.com]

Elicitation of purposive text was carefully designed to sample across the range of experiences found within the 10 primary-care clinics in the organization (for more information on sampling design and rationale, see Tomoaia-Cotisel (2018)). Individual semistructured interviews were conducted with health-services stakeholders – each focusing on the individual's experiences as they wrestled with their own primary-care transformation journey. The

DOI: 10.1002/sdr

interviews from a portion of these clinics ($n = 5$ clinics) were used for model development (i.e. development data set), and the rest were set aside for model validation (i.e. validation data set). RIQ analysis uses one clinic from this validation data set which included interviews of the center manager, medical director, and nurse manager, as well as one clinical staff member and one clinician from each care team.

The RIQ method (shown with a star in Figure 1) was one of many validation methods introduced in the parent study. The RIQ method aims to subject the contemporary dynamic hypothesis that is the CLD generated from the first portion of interviews to additional purposive text. It does this to confirm, disconfirm, and revise that CLD (and thus to revise the dynamic hypothesis). The scope of this article focuses on this specific role of RIQ analysis. However, it is worth noting that the parent study uses the RIQ method beyond this validation purpose, and additional methods such as saturation analysis and stakeholder dialog were also used to evaluate the CLD generated (Tomoaia-Cotisel, 2018).

Once sufficient confidence was obtained, the revised CLD was used as the blueprint for the building of the simulation model. The validation data set was used again later to validate the simulation model. After undergoing the additional set of standard system dynamics validation tests (Barlas, 1996; Lane, 1995; Sterman, 2000), the simulation model was used for policy analysis and to facilitate stakeholder dialogues to discuss the results obtained and their policy implications. Finally, a series of reflections were undertaken to contribute to relevant theory and to methods development.

Methods for rigorously interpreted quotation analysis

In this section, we describe how our interview data was analyzed using the RIQ method to extract evidence that either confirms or disconfirms the elements of an existing CLD (these include variables, links, delays, and feedback loops). We present *the logic* for RIQ analysis using a table. After each table, we present a figure that visualizes what the table accomplishes in terms of CLD evaluation, such as what is verified and what revisions are proposed.

For the purpose of illustrating the method, we present four quotations from primary-care clinicians. We define a quotation as a chunk of text expressing a cohesive story. The first quotation presents the simplest example of the RIQ method, where the data confirms some elements in the existing CLD while also suggesting a need for revisions. The second quotation is slightly more complex than the first. It not only provides further support for the confirmation and structural revision identified in the first quotation, but it also illustrates the confirmation of the perception of a delay and a feedback loop identified in the CLD. The third quotation presents a case, referred

DOI: 10.1002/sdr

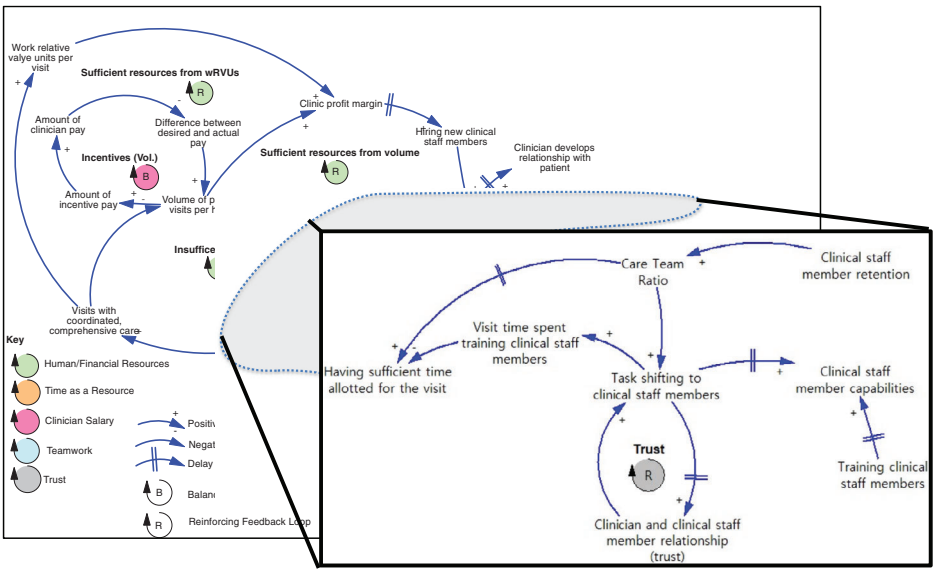

Fig. 2. Pre-RIQ CLD with Zoomed-in Portion for Illustration of the RIQ Method. Arrows that are straight in the zoomed-in box are not in loops, and ones that are curvy are in loops when viewing the entire diagram. This is done so that links in the inset CLD follow best practice (Sterman, 2000, 2018a, 2018b). This figure is adapted from the original version found in Tomoaia-Cotisel (2018) [Color figure can be viewed at wileyonlinelibrary.com]

to as an implicit mention, where a simple causal structure identified in the quotation is matched with more complex, detailed structure in the existing CLD. By assuming that what is implied in the third quotation does not refute the existing CLD's structure and by documenting this assumption, we capture the amount of evidence for each link in the existing CLD. The fourth quotation is an example illustrating how structures evaluated using the RIQ method are often complexly embedded within stories that meander through different parts of the system.

The quotations presented here relate to a small section of the CLD generated from the initial conceptualization process, referred to as the "pre-RIQ" CLD (see CLD inset in Figure 2). The simulation model[viii] that was ultimately produced also has a section that corresponds to this one. This particular section was chosen for illustration of the RIQ method for two main reasons: (i) it shows a core mechanism of the simulation model developed in the parent study, *and* (ii) this section of the CLD was substantively revised using the RIQ method before serving as a blueprint for that simulation model. This section of the CLD has a reinforcing loop labeled "Trust," where the more trust clinicians have in their clinical staff members, the more tasks will be shifted to the clinical staff members, providing more opportunities for the clinicians to build trust in their clinical staff members. With task shifting, clinical staff members can further develop their capabilities. Formal training

[viii]This simulation model was built using a revised CLD as a blueprint, as described in the "Research Context for Method Development" above. See Figure 8 in "Results of CLD refinement" section below for discussion of the section of the simulation model that was informed by the section of the CLD discussed in this article.

DOI: 10.1002/sdr

can also increase these capabilities. Having a higher care-team ratio (clinical staff members to clinician ratio) allows for an increased amount of task shifting, and, over time, once training is completed, it also results in having enough time to do what is needed inside the visit time. Turnover would reduce this ratio.

How rigorously interpreted quotation analysis works

To illustrate the RIQ method, we use interview data from primary-care clinicians. Each interview is labeled with a data source code, and for these clinicians, the codes are CL01, CL03 and CL04. Like the other primary-care clinicians, these clinicians work on teams with other clinicians and clinical staff members. Each shared their perspective on the transformation efforts in primary-care service delivery.

In the quotation below, CL04 describes why they do not delegate more tasks to clinical staff members on the team. This quotation went through a minor clean-up and is referred to as a *truncated* quotation. Duplicate words or conversation fillers are removed to enhance readability, and personal identifiers are removed to protect identity.

> If I send out a note to call this patient … sometimes I will go back and look … days later and no one's called them. That worries me quite a bit. Sometimes I feel like I do more things than I probably should just because I would just rather just do it myself and get it done correctly… I try and stick with basic things for the [clinical staff members] to do and if [it is at all] complicated … then I just do it… I guess I can mostly rely on people but some things I don't trust them with and I would rather just do it myself.

From the quotation, we can infer that while it would be preferred to shift tasks to clinical staff members, when clinical staff members fail to perform these tasks, the clinician loses trust in them and avoids shifting tasks. This quotation is now ready for RIQ analysis.

Table 1 shows the basic set up of RIQ analysis. The left side shows the quotation, and the right side shows how the quotation is interpreted in relation to the CLD. More specifically, the upper-right side breaks down the quotation into small phrases from which each phrase is interpreted in relation to the CLD elements (i.e. variables, causal relationships including arrow direction and polarity, delays, feedback loops). The lower-right side summarizes the coder's interpretation of the quotation into a causal statement that can be tied to the CLD.

The quotation is entered into the left side of the RIQ table. When entered, we keep track of who said the quotation (CL04) and where this quotation appears in the interview (quotation number 30). In the quotation entered, phrases that are relevant to an element in the CLD are underlined for

DOI: 10.1002/sdr

Table 1. RIQ example: analysis resulting in adding a new link

Quotation	Interpretation	
CL04-30) "If I send out a note to call this patient … sometimes I will go back and look … days later and no one's called them. That worries me quite a bit. "Sometimes I feel like I do more things than I probably should just because I would just rather just do it myself and get it done correctly… I try and stick with basic things for the [clinical staff members] to do and if [it is at all] complicated … then I just do it… "I guess I can mostly rely on people but some things I do not trust them with and I would rather just do it myself." (110/188)	Phrases • no one's called them • That worries me quite a bit • I would just rather just do it myself and get it done correctly • I guess I can mostly rely on people but some things I do not trust them with • just do it myself • If I send out a note to call this patient • I do more things than I probably should • stick with basic things for the [clinical staff members] to do and if [it is at all] complicated … then I just do it • just do it myself	CLD Elements Clinical staff member capabilities Clinician and clinical staff member relationship (trust) Task shifting to clinical staff members
	Causal Chain Clinical staff member capabilities →+ Clinician and clinical staff member relationship (trust)→+ Task Shifting to clinical staff members	
	Coder's Interpretive Notes The clinician makes the decision to shift tasks by considering how much they trust clinical staff members. When clinical staff member performance on assigned tasks is poor, then trust is low. There is a basic level of tasks that clinical staff members can do which do not require the clinician's trust.	

interpretation. They would be double underlined if used more than once in our interpretation. Finally, we keep track of the current word count of the truncated quotation as well as the total word count of the untruncated, original quotation. This is provided for transparency. Furthermore, if a truncated quotation's interpretation raises a question, the word counts may flag that it might be worth reviewing the original transcript to see what has been removed in the truncation.

The right side of the RIQ table shows the interpretation of the quotation in four subsections: *Phrases*, *CLD Elements*, *Causal Chain*, and *Coder's Interpretive Notes*. In the *Phrases* section, we copy and paste the phrases from the quotation that are relevant to the CLD being assessed. A phrase may confirm variables already existing in the CLD. Matching variables are listed next

DOI: 10.1002/sdr

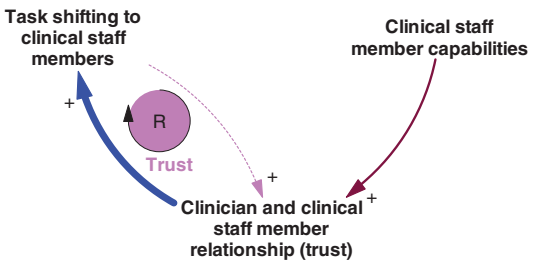

Fig. 3. RIQ Example: CLD Visualizing RIQ in Table 1 [Color figure can be viewed at wileyonlinelibrary.com]

to the phrase in the *CLD Elements* section. If the phrase mentions new CLD elements or contradicts the existing CLD elements, the revised elements are listed and marked to show a revision (highlighted yellow). We have found it helpful to list the phrases and CLD elements in order of the causal chain, even if the conversation starts somewhere downstream of the identified causal chain. This aids the reader of the RIQ table to better follow what is being documented.

In the *Causal Chain* section, the CLD elements and the causal relationship (s) among them are summarized in the form of the causal link (s) with polarity and delays. If the causal chain includes a revision, then it is marked to show the revision (highlighted yellow). This chain is the result of the *rigorous interpretation* of the given quotation. Later, we also show how to organize an RIQ table if a quotation identifies multiple causal chains (See Table 4). In the *Coder's Interpretive Notes* section, the coder writes their reasoning for the causality that is documented – it is crucial that the coder grounds their notes in what is shared in the quotation rather than just restating the causal links in sentence form.

The above quotation from clinician CL04 was able to confirm existing elements in the CLD and proposed a new causal link to be added to the CLD. Below we summarize the result and visualize them in Figure 3:

- The quotation verified the following elements existing in the pre-RIQ CLD:
  - three variables: Task shifting to clinical staff members, Clinical-staff-member capabilities, and Clinician and clinical-staff-member relationship (Trust)
  - the positive relationship from Clinician and clinical-staff-member relationship (trust) to Task shifting to clinical staff members [bolded in **blue** in Figure 3]
- The quotation proposed a revision to CLD:
  - the addition of the positive relationship from Clinical-staff-member capabilities to Clinician and clinical-staff-member relationship (Trust) [shown in **maroon** in Figure 3]
- The quotation did not verify the following pre-RIQ CLD elements:

DOI: 10.1002/sdr

- the positive relationship from Task shifting to clinical staff members to Clinician and clinical-staff-member relationship (trust) [dashed **lavender** in Figure 3]
- the *Trust* loop [shown in **lavender** in Figure 3]

How delays and feedback loops are identified

The first example illustrated how variables and causal relationships are extracted from a quotation. In this section, we show the process for identifying a delay and a feedback loop using RIQ analysis. For simplicity of explanation, we continue to use interview data focusing on the same region of the CLD, from another clinician (CL01).

In the same way that a phrase may confirm a variable, a phrase can also confirm a delay or a feedback loop already existing in the CLD. Matching feedback loops and delays are listed next to the phrase (s) in the *CLD Elements* section. These CLD elements are entered in parentheses to distinguish them from the variables found in the CLD. They are also listed before variables in this section since they highlight key dynamics existing in the quotation.

As shown in the Quotation section of Table 2, CL01 explains what they do when clinical staff members do not exhibit the capabilities needed to perform shifted tasks. The clinician notices there is a learning curve. After some time, capabilities would improve, and the clinician would shift more tasks. Notice that in addition to the CLD verification and restructuring proposed in the first example, this quotation also explicitly identifies a delay and a feedback loop.

To be coded as the CLD element of "(Feedback Loop)," the phrase *does not need to* use the word "feedback" or to give a description of communicating opinions (colloquially referred to as providing someone feedback). The phrase *does need to* indicate the presence of a causal loop. In this quotation, the clinician describes working to develop capabilities as they shift tasks to clinical staff members and part of that capability development is communicating about how shifted tasks are being performed (providing that feedback when performance is not up to the standard). As those capabilities are demonstrated, trust is enhanced and further task shifting takes place. This is the "(Feedback Loop)" that the clinician is explicitly identifying in this quote.

RIQ analysis pulls out these phrases and labels them accordingly.

The above quotation from CL01 was able to confirm existing elements in the CLD. Some of these elements were verified in the first example, and they were verified again here. One CLD element was newly verified, with ample evidence that CL01 perceives the delay in the link. The quotation supports the new causal relationship proposed in the first example and also shows

DOI: 10.1002/sdr

Table 2. RIQ example: analysis verifying delay & feedback loop

Quotation	Interpretation	
CL01-31) "I certainly like to give feedback when something wasn't done up to my standards and normally that helps… I am a stickler about always [doing x, for example,] … and usually, with new [clinical staff members], they never [do x]. And then, after a couple of times, that usually changes… If I can find them … [then] I do it face to face. [It] normally [takes] a day or two… "[In this department] we all like things done a different way… So usually it takes [MAs] a little bit … to get used to that. But normally they learn." (99/188)	Phrases • after a couple of times • It normally takes a day or two • it takes [clinical staff members] a little bit • I certainly like to give feedback when something wasn't done up to my standards and normally that helps • something wasn't done up to my standards • with new [clinical staff members] they never do x … that usually changes • they learn • I certainly like to give feedback • If I can find them, then I do it face to face • I am a stickler about • we all like things done a different way	CLD Elements (Delay) (Feedback Loop) Clinical staff member capabilities Clinician and clinical staff member relationship (trust) Task shifting to clinical staff members
	Causal Chain Clinical staff member capabilities →+ Clinician and clinical staff member relationship (trust)→+ Task Shifting to clinical staff members --||→+ Clinical staff member capabilities Coder's Interpretive Notes Capabilities development takes time. This includes 1) the time for the clinicians to find the clinical staff member to bring up a deficiency and 2) the time for the clinical staff member to learn to get it right.	

evidence of CL01's perception of the new feedback loop that this added link creates. Below, we summarize the results and visualize them in Figure 4:

- The quotation verified the following elements existing in the pre-RIQ CLD:
  - the three variables: Task shifting to clinical staff members, Clinical-staff-member capabilities, and Clinician and clinical-staff-member relationship (Trust)
  - the positive relationship from Clinician and clinical-staff-member relationship (Trust) to Task shifting to clinical staff members [bolded in **blue** in Figure 4]

DOI: 10.1002/sdr

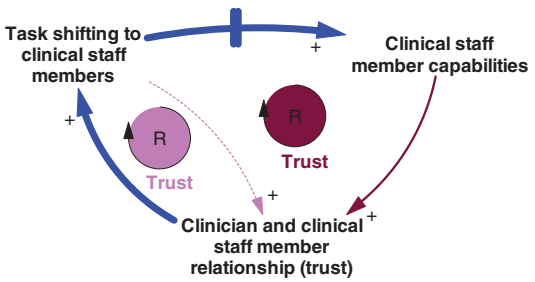

Fig. 4. RIQ Example: CLD Visualizing RIQ in Table 2 [Color figure can be viewed at wileyonlinelibrary.com]

  - the positive relationship with a delay from Task shifting to clinical staff members to Clinical-staff-member capabilities [bolded in **blue**)]
- The quotation proposed a revision to the pre-RIQ CLD:
  - the positive relationship proposed from Clinical-staff-member capabilities to Clinician and clinical-staff-member relationship (Trust) [shown in **maroon**)];
  - addition of a new *Trust* loop [shown in **maroon**)]
- The quotation could not verify the following pre-RIQ CLD elements:
  - the positive relationship from Task shifting to clinical staff members to Clinician and clinical-staff-member relationship (trust) [dashed **lavender**]
  - the original *Trust* loop [shown in **lavender**]

How implicit structures are identified

It is not uncommon for a participant's description to skip steps in a causal chain. In these cases, it is useful to evaluate if a more complex mechanism existing in the CLD can suitably capture that experience, or if in fact the participant's description is a distinct mechanism that should be added. In this section, we illustrate how variables and links on the CLD may sometimes only be implicitly mentioned in a quotation, and how we document such variables in RIQ analysis. For simplicity of explanation, we continue to use interview data focusing on the same region of the CLD, from another clinician (CL03).

If we look at the pre-RIQ CLD in Figure 2, there are two causal paths specifying both a short-term and a long-term effect of having a higher staff ratio:

- short-term effect of more training needs: Care-team ratio →+ Task shifting to clinical staff members →+ visit time spent training clinical staff members →- Having sufficient time allotted for the visit
- long-term effect of more time with the patient: Care-team ratio --||→+ Having sufficient time allotted for the visit

DOI: 10.1002/sdr

As shown in the Quotation section of Table 3, CL03 describes what they experience when clinicians shift tasks from themselves to their current team of clinical staff members: those staff members do not have enough time to get shifted tasks done in addition to their preexisting tasks so, instead, clinical staff members just do not do those tasks. CL03 hypothesizes: if they had an additional person specifically who would have time to do these tasks that clinicians are shifting, then the task shifting would be possible and valued. In this context, *Task shifting to clinical staff members* impedes *Having sufficient time allotted for the visit*, because it makes clinical staff members too busy to perform all of the tasks the clinician has assigned – that is:

- Task shifting to clinical staff members →- Having sufficient time allotted for the visit

In such a case, the negative relationship may be added to the pre-RIQ CLD as a new causal path. However, it is also plausible to assume[ix] that CL03 is implying the following without explicit mention of the mechanisms:

A. in the short term, clinical staff members cannot keep up with all the tasks shifted to them, because time is taken up participating in on-the-job training. As a result, patient visits are delayed (hence we are assuming an implicit mention of training time in this specific quotation) and
B. in the long term, when on-the-job training is no longer needed, having staff members designated for shifted tasks allows for sufficient time for patient visits (as described in other interviews and visualized on the CLD via other links with delays).

Based on this assumption, we determined the quotation does not require modification of the CLD as it is captured by the existing causal paths. In the Table 3, we mark this as an implicit mention by placing a "0" in the *Phrases* section for the missing intermediary variable. In the *Causal Chain* section, the part of the causal chain that is mentioned implicitly is highlighted in gray.

In this quotation, the concept of capacity constraint is also implied, and it led to revising the name of the variable *Care-Team Ratio* to *Clinical-Staff-Member Capacity* for further clarification. *Care-Team Ratio* was the idiosyncratic term used within this case to describe the ratio of clinical staff members to clinicians; however, when applying it to a given situation, participants exclusively talked about changes to the numerator of this ratio (the number of clinical staff members).

[ix]In the parent study's CLD, there were multiple causal paths that could represent this particular quote. Here we picked out one possible causal path as an example. Adding another causal path to this table was avoided for parsimony.

DOI: 10.1002/sdr

Table 3. RIQ example: analysis including implicit structures

Quotation	Interpretation	
CL03-21) "The potential is there to … make a huge impact on how we take care of our patients… in terms of the quality measures. … If we had a designated person [again]… going through and calling every single patient with diabetes … to come in and get their labs … and reminding … going through my schedule … and [saying] 'these 5 patients need [x]' and giving the [clinical staff member] a little note saying, 'could you please order these [tests]'… That's what the [clinical staff members] are supposed to be doing but they are doing so many other things when they are going in there, when that … reminder pops up [on their screen during the visit], they are just like "oh, ok [the doctor] will do it…" (129/300)	Phrases • a designated person • going through and calling every single patient with diabetes … to come in and get their labs • reminding… going through my schedule… [saying] 'these 5 patients need [x]' • giving the [clinical staff member] a little note saying, 'could you please order these [tests]' • they are doing so many other things when they are going in there • what the [clinical staff members] are supposed to be doing 0 • That's what the [clinical staff members] are supposed to be doing but • when that … reminder pops up [on their screen during the visit], they are just like "oh, ok [the doctor] will do it…"	CLD Elements Clinical Staff Member Capacity Task shifting to clinical staff members Visit time spent training clinical staff members Having sufficient time allotted for the visit
	Causal Chain Clinical Staff Member Capacity →+ Task shifting to clinical staff members →+ Visit time spent training clinical staff members →- Having sufficient time allotted for the visit Coder's Interpretive Notes Staffing on this team is constrained such that clinical staff members do not have the time needed to complete all of the tasks that they are asked to complete when there is task shifting. They leave some tasks undone and they assume the clinician will get to (at least some of) these tasks since they do not have the time to do so. The clinician indicates that more staffing (such as a designated person to do some of the shiftable tasks) is needed for visits to have enough time. Causal Chain Clinical Staff Member Capacity --||→+ Having sufficient time allotted for the visit Coder's Interpretive Notes This quotation identifies these variables and the link between them: having an additional team member would allow the team to have enough time in the visit; but it does not mention the delay.	

DOI: 10.1002/sdr

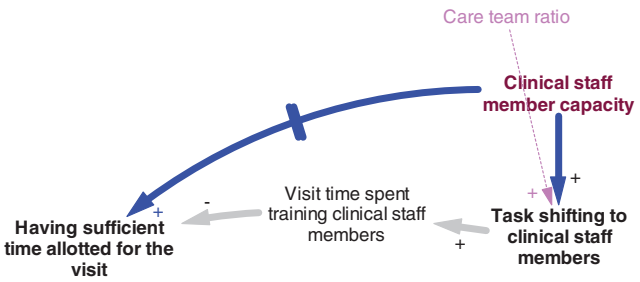

Fig. 5. RIQ Example: CLD Visualizing RIQ in Table 3 [Color figure can be viewed at wileyonlinelibrary.com]

Above quotation from CL03 led to the following results, and they are visualized in Figure 5:

- The quotation verified the following elements existing in the pre-RIQ CLD:
  - three variables: Clinical-Staff-Member Capacity, Task shifting to clinical staff members, Having sufficient time allotted for the visit [**bolded** in Figure 5]
  - the positive relationship from Clinical-Staff-Member Capacity to Task shifting to clinical staff members [bolded in **blue**)]
- The quotation includes a variable in the CLD with a name change: *Care-team ratio* is changed to *Clinical-Staff-Member Capacity* [variable shown in **maroon** replaces **lavender** variable].
- The following CLD elements are implicitly supported by the quotation:
  - one variable: Visit time spent training clinical staff members
  - the positive relationship from Task shifting to clinical staff members to Visit time spent training clinical staff members [link shown in **gray**]
  - the negative relationship from Visit time spent training clinical staff members to Having sufficient time allotted for the visit [link shown in **gray**]

How structures are identified from meandering conversations

In a typical analysis of interview data, the RIQ method can be a more complex process than what is presented with the previous quotation examples. In our fourth example, we illustrate the degree of complexity one might come across when interpreting semistructured data that flows as conversations often do – in a meandering way. This quotation comes from an earlier part of the interview with clinician CL04. In this quotation, the clinician discusses facilitators and barriers to developing a high level of trust that enables task shifting.

DOI: 10.1002/sdr

Table 4. RIQ Example: Analysis of a Meandering Conversations

Quotation	Interpretation	
CL04-29) "But when I was at [another clinic] … it was kind of working… some [clinical staff members] worked better with certain providers than others so, that was kind of nice because certain [clinical staff members] you just clicked with, and you could get things done a lot quicker. They knew how to … [do] things … while we were talking, and they could start your note for you… If we could train all the [clinical staff members]to do that and have enough staffing to do it, then that would be great. But it seems like that takes a lot of work because some of them aren't trained or willing to learn… "Staff turnover … is a big issue … we have changed … the [clinical staff members] … [None of the 7 on the team have] worked here [more than 2 years] "I have to tell [clinical staff members] little things about what I would like to have done [for x, prepare y] … I just feel like I continually have to tell people that. So, it does take a lot more time and patience… "Yeah, they get plenty of training, but they do not get training on specifically what I like… because I just have to tell them constantly." (204/444)	Phrases • it seems like that takes a lot of work • it does take a lot more time and patience • get things done a lot quicker • they knew how • some of them aren't trained • some [clinical staff members] worked better with certain providers than others so • certain [clinical staff members] you just clicked with • willing to learn • I just feel like I continually have to tell people • I just have to tell them constantly • get things done • do things while we were talking, and they could start your note for you • to do it Causal Chain Clinical staff member capabilities →+ Clinician and clinical staff member relationship (trust)→+ Task shifting to clinical staff members --||→+ Clinical staff member capabilities Coder's Interpretive Notes The choice to place trust in a clinical staff member is sensitive to the clinician's assessment of the clinical staff members ability to learn. Without passing this test, CL04 will not initiate task shifting, even with experience in the model and substantial buy-in. CL04 admits the true time required to train is unknown, but it seems long. Because of the high rate of turnover, CL04 experiences it to be infinitely long.	CLD Elements (Delay) Clinical staff member capabilities Clinician and clinical staff member relationship (trust) Task shifting to clinical staff members
	Phrases • If we could train all the [clinical staff members] • they get plenty of training Causal Chain Off-the-job clinical staff member training --||→+ Clinical staff member capabilities →+ Clinician and clinical staff member relationship (trust) Coder's Interpretive Notes There are two types of training: training done by the clinician on-the-job, and training done by the organization. The clinician sees the organization led training as an easier, faster way to develop capabilities such that clinical staff members can be trusted to perform at a higher level of task-shifting. Otherwise, task-shifting feels burdensome because it requires so much time spent in on-the-job training.	CLD Elements Off-the-job clinical staff member training

(Continues)

DOI: 10.1002/sdr

Table 4. Continued

Quotation	Interpretation	
	Phrases	CLD Elements
	• I have to tell [clinical staff members] little things about what I would like to have done [for x, prepare y] … I just feel like I continually have to tell people	Visit time spent training clinical staff members
	• but they do not get training on specifically what I like	
	Causal Chain	
	Task shifting to clinical staff members →+ Visit time spent training clinical staff members	
	Coder's Interpretive Notes	
	The clinician feels compelled to explain their preferences for how each task that is shifted should be performed. This on-the-job training takes time and patience.	
	Phrases	CLD Elements
	• Staff turnover … we have changed … the [clinical staff members]	Clinical staff member retention
	• have enough staffing	Clinical Staff Member Capacity
	Causal Chain	
	Clinical staff member retention →+ Clinical Staff Member Capacity →+ Task shifting to clinical staff members	
	Coder's Interpretive Notes	
	The choice to shift tasks is sensitive to the amount of staffing, and to the length of tenure of clinical staff members. High turnover has been a big issue for this team.	

Unlike the previous examples, Table 4 has a long quotation which results in four causal chains. We are not able to split the quotation into four shorter quotations, each with one causal chain, because the main subject of the quotation is linked to different causal chains scattered throughout the quotation. In this case, keeping the subject matter together makes the interpretation more useful. Quotations often meander through dispersed ideas, as brief bits of verbal expression provide only a glimpse of one's underlying mental model. In such a situation, the coder must tease out CLD elements from the data and make causal relationships explicit. With multiple causal chains in a quotation, each distinct chain is given a separate row and interpreted one at a time in the same table. The Coder's Interpretive Notes section is particularly important for such quotations as it requires the coder to stop at each individual causal chain, reflect on the interpretations made, and document the reasoning.

The RIQ shown in Table 4 verified many of the same CLD elements from the previous examples. It also verifies other variables found in the pre-RIQ

DOI: 10.1002/sdr

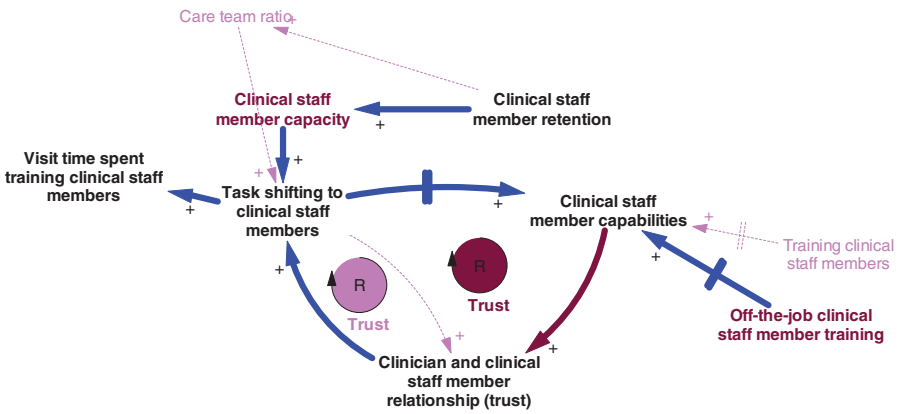

Fig. 6. RIQ Example: CLD Visualizing RIQ in Table 4 [Color figure can be viewed at wileyonlinelibrary.com]

CLD. Furthermore, this quotation supports changing the name of the variable *Training clinical staff members* to *Off-the-job clinical-staff-member training*. The quotation clarifies that on-the-job training involves clinicians spending visit time providing this training and that one-time trainings done by the organization might also be useful. In so doing, this quotation supports distinguishing between these two types of training. The variable name is changed to include an indication that it is referring to organizationally provided training.

New contributions from this quotation are summarized below and they are visualized in Figure 6.

- The quotation verified the two additional variables from the pre-RIQ CLD: *Clinical-staff-members retention* and *Visit time spent training clinical staff members*.
- The quotation includes a variable in the CLD with a proposed name change: *Training clinical staff members* is changed to *Off-the-job clinical-staff-member training* [variable shown in **maroon** replaces **lavender** variable]

Results of CLD refinement with rigorously interpreted quotation analysis

Substantive changes to the "Pre-RIQ" CLD

In comparing our "pre-RIQ" and "post-RIQ" CLD, we see that substantive changes were made to the initial understanding of this section of the diagram. For instance, this process led to the recognition that capability

DOI: 10.1002/sdr

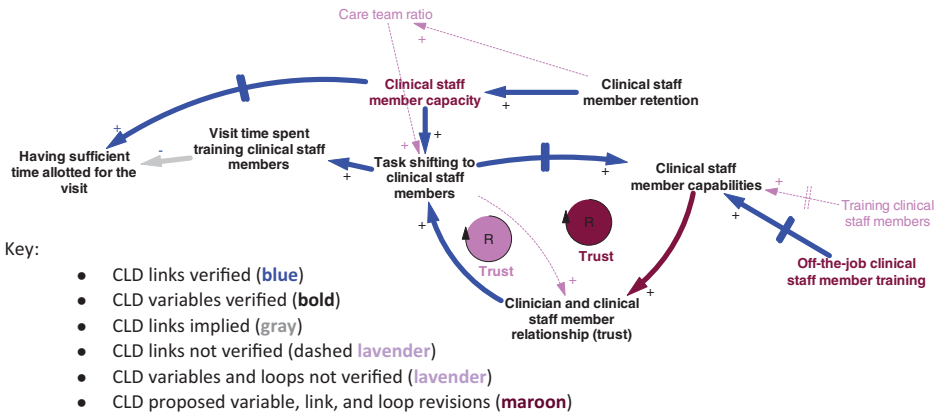

Fig. 7. Post-RIQ CLD Visualizing Results of RIQ Analysis in this Article. Key: CLD links verified (**blue**). CLD variables verified (**bold**). CLD links implied (**gray**). CLD links not verified (dashed **lavender**). CLD variables and loops not verified (**lavender**). CLD proposed variable, link, and loop revisions (**maroon**) [Color figure can be viewed at wileyonlinelibrary.com]

Table 5. Results of RIQ analysis – comparing pre-RIQ and post-RIQ CLDs

CLD Elements	Pre-RIQ CLD Total Counts	RIQ Analysis				Post-RIQ CLD
		Verified	Not Verified	Dropped	Added	Total Counts
Variables	22	20[a]	1	1	0	21
Relationships	31	24	3	4	3	30
Delays	11	7	1	3	0	8
Distinct Feedback Loops	10	7[a]	1	2	2	10

[a]Four variables were renamed during this process. Two loops were renamed.
This table is based upon results reported in Table F.16 in (Tomoaia-Cotisel, 2018).
Pre-RIQ count = Verified + Not Verified + Dropped.
Post-RIQ count = Verified + Not Verified + Added = Total Counts Prior to RIQ + Added – Dropped.

development is *endogenous* to the feedback process of developing trust – a structure that was highly impactful in the simulation model. Figure 7 is the inset section of the "post-RIQ" CLD that was generated as a result of the confirmations and revisions made from the RIQ analysis. This figure visually illustrates the CLD elements that were verified and those that were modified during RIQ analysis. Table 5 provides counts of the CLD elements existing in the entire "pre-RIQ" CLD that were verified and the modifications made to create the "post-RIQ" CLD. It is important to note that the magnitude of changes reported (i.e. elements dropped and/or added) is not the focus of Table 5. This table is provided to suggest that the RIQ analysis improved the quality of the product.

The revised *Trust* loop (shown in Figure 7 above) that resulted from the RIQ process became a powerful feedback mechanism that is core to the

DOI: 10.1002/sdr

simulation model's structure and its policy insights.[x] See Figure A2 in the online supporting information for a diagram of the structure of the *Trust* loop in the simulation model.

Effort expended in rigorously interpreted quotation analysis

Various factors determine the resources required to implement RIQ analysis. In the case of the research project discussed in this article, we identified and conducted RIQ analysis for, on average, 24 separate quotations per interview (where each interview had 20 pages of single-spaced text on average). These interviews and quotations varied widely in their length, quality, and the degree of relevance to the CLD structure being evaluated. In our case, an experienced coder spent from 5 minutes to an hour on one quotation.

In general, the following factors increased the time needed to analyze a quotation:

- the length of the quotation,
- whether the quotation contains proposed restructuring,
- how many variables the quotation relates to,
- the breadth of variety of the situations discussed in the quotation,
- how much meandering is happening in the quotation, and
- how much extra text (noncausal language) is there to be truncated.

In addition, several coder characteristics facilitated RIQ analysis:

- the coder's skill at diagramming (i.e. identifying causal language in text, following CLD diagramming conventions), and
- the coder's familiarity with the problem context (i.e. with conceptual information, descriptions of concepts that facilitated formulating variable names).

Finally, the quality of the initial CLD was also an important factor in determining how much effort RIQ analysis might require. Revisions due to disconfirmation took more time to figure out than quotations that confirmed existing structure.

A broader reflection on the return on investment of rigorously interpreted quotation analysis

We found that RIQ analysis resulted in benefits that outweighed the effort needed. The process of CLD verification and modification resulted in

[x]The parent study (Tomoaia-Cotisel, 2018) provides the operational details of the *Trust* Loop in the final simulation model.

DOI: 10.1002/sdr

insights that were not available in the earlier conceptualization phase. Unlike the inductive process of analyzing data in the initial CLD generation, RIQ analysis is a deductive process: whereas the initial CLD generation involved a discovery of emerging constructs from raw data and capturing patterns in the form of CLD, the process of RIQ analysis required one to look for elements in the data that could verify or modify the existing CLD. This shift generated valuable new information. It permitted revisiting each element in the CLD, triangulating evidence regarding the element, and engaging in deep reflection about its role in the CLD.

We also believe that, while RIQ analysis required additional time in the conceptualization phase, it saved more than enough time in development of a simulation model. Along with the initial pre-RIQ CLD, tentative simulation models were also developed early in the project. It was challenging to balance the need to add structure only as necessary with the need to reconcile differences across participants. RIQ analysis resolved these concerns to a great extent as: (i) revisions to the initial CLD-clarified understanding of the links surrounding important variables like task shifting and (ii) much of the structure was repeatedly confirmed (Tomoaia-Cotisel, 2018).[xi] As long as it is focused on the problem and uses purposive text data, RIQ analysis helps to make revisions that improve the CLD's quality. This is the case whether revisions are the result of differences in context or from the modeler gaining a new understanding about the original context from new data. These revisions and supporting evidence from RIQ analysis bolstered confidence in the suitability of the CLD's variables and links and the adequacy of the model boundary.

This confidence helped the modeler to focus on the problem rather than to "model the system" (Sterman, 2000, p. 79). RIQ analysis helped the modeler to abandon several preliminary simulation model structures that had led to dead ends. RIQ analysis helped the modeler to feel confident that the post-RIQ CLD was an adequate endogenous theory of the part of the system causing the problem and the timeframe over which important changes were occurring. This confidence allowed the modeler to distinguish between the prior model structures which were unintentionally pushing toward modeling "the system" and the ones that needed to be in the model.

Each RIQ table imitates a *conversation* between a modeler and a stakeholder about the CLD. The table shows *how the modeler is listening* to a unique perspective based in a specific experience. Assigning phrases to variables and writing interpretations about causal chains documents *how the modeler is interpreting* similarities and differences in perspective across individuals, coming from their varied experiences. Every quotation is an opportunity to question the CLD and to figure out *how to reconcile any differences* between the perspective in the quotation and the structure in the

[xi]See Section 3.5.4.2 and Section 2.4.5.4.2.

DOI: 10.1002/sdr

CLD. Ideally, the CLD should be able to describe the important dynamic structures across sites. Focusing conceptualization on the interpretation of one discrete experience improves the CLD's quality. Performing RIQ analysis many times also provides information that can be used to assess quality more holistically; for example, via within-case and cross-case comparisons. When a simulation model based in the post-RIQ CLD is developed, this information makes possible further opportunities for improving quality.

Finally, the RIQ tables generated as part of the analysis facilitated discussions with research stakeholders that were grounded in the field data. These included discussions with the dissertation committee members (each with diverse expertise), dissertation examiners, and experts in health-services research. In so doing, the RIQ tables facilitated CLD evaluation and the building of shared understanding and confidence in the content of the CLD and, ultimately, in the simulation model and insights derived from it.

Conclusion and discussion

In this article, we present the RIQ method for comparing a CLD to qualitative purposive text data. Using RIQ analysis to evaluate a diagram during the conceptualization process has many benefits including: (i) enabling a systematic and transparent approach to dissecting reliable qualitative data which communicates experience in natural language, (ii) documenting that process, and (iii) documenting evidence that confirms prior conceptualization and evidence that disconfirms and suggests revisions to prior conceptualization (thus, backward looking, allowing for the tracing CLD elements to evidence). Presenting RIQ results alongside a CLD allows researchers to share supporting data in the natural language of the data source along with documentation for how that data was interpreted to arrive at the results obtained (thus, RIQ analysis is forward looking as it allows for reproducibility and further disconfirmation). We believe RIQ analysis offer an important avenue for building confidence in the usefulness of the findings and recommendations.

More specifically, RIQ analysis offers scientific and practical value for different project contexts and stakeholders. For a team of researchers, the RIQ table's systematic way of interpreting data and documenting the interpretive process can enhance communication in the team. It reduces variability in interpretation among team members and would make it easier to pick up where previous system dynamics modeling efforts left off. For clients, RIQ analysis ensures their voices are formally captured in the resulting model. By being able to trace their own input in the model, clients experience a greater sense of ownership of the overall modeling process. This is critical for the clients to trust the final model outcomes as it will result in a higher likelihood for model recommendations to be implemented. Broadly

DOI: 10.1002/sdr

speaking, RIQ analysis allows qualitative modeling work to be evaluated, replicated, and disconfirmed by peers. This allows CLD structures to be evaluated in diverse situations.

We believe our article will speak to those new to system dynamics (herein novices) and system dynamics experts alike. RIQ analysis relies on thematic analysis (broadly defined as coding chunks of text data) and CLDs, both of which are widely used for their accessibility. Therefore, we expect RIQ analysis to be easily implemented by novices and experts alike. (In fact, along with supporting empirical rigor, ease of use was a major motivation for the full suite of methods developed in Tomoaia-Cotisel (2018).)

For novices, RIQ analysis can be especially useful for *building one's own confidence* that the conceptualization is rigorous. Returning to the elephant story, RIQ analysis can mitigate (i) coming up with spears and snakes (i.e. misinterpretation of evidence) and (ii) wasted time debating whether one is seeing parts of an elephant or other things (i.e. searching for interpretations without relying on evidence). Furthermore, for those planning to use a CLD as a blueprint for simulation, RIQ analysis provides a rigorous understanding of what ought to be considered as they embark upon formalizing their simulation model.

For system dynamics experts (researchers, consultants, and practitioners), RIQ analysis can be especially useful for building others' confidence that the conceptualization is rigorous. Again, returning to the elephant story, RIQ analysis can mitigate acting as the blind people who each continues to see a spear or a snake as they are presented with more evidence (i.e. bringing and holding on to one's own interpretation). We expect system dynamics experts to find that the logic of RIQ analysis (herein presented as the RIQ table) is robust and flexible for use in various applications.

More broadly, we feel that the RIQ method contributes to responding effectively to Richardson's call (1999) for principles of qualitative system dynamics. Specifically, we feel that the RIQ method contributes to building the "more robust catalogue of qualitative 'methods'" (Richardson, 1999) (p.441) for improving the quality of CLDs. We also believe that the RIQ method contributes to Sterman's (2018c) call for innovating methods that facilitate rigorous evidence-based testing of diagrams and for improving projects' empirical rigor and policy impact (Sterman, 2018c). Finally, we believe that our article, and the RIQ method specifically, contributes to expanding system dynamics' engagement with social science beyond those approaches described previously (cf. Kopainsky and Luna-Reyes, 2008, and Luna-Reyes and Andersen, 2003).

Indeed, calls to use empirically rigorous qualitative techniques for causal identification and inference are being made by scholars who are increasingly concerned about how weakness on this point hampers policy impact (cf. Miller *et al.*, 2013, and Reiss, 2009). More rigorous conceptualization may not prove beyond a doubt that a causal link is always real or that

DOI: 10.1002/sdr

manipulating a cause will invariably guarantee the desired effect, but, when using a tool like a CLD, the RIQ method can improve understanding of dynamically complex structure across stakeholder groups in a way that reflects their varying experiences with a problem. This provides a salve for narrow mental models that we believe is desperately needed in the shared battle against reductionism.

Limitations

This article presents examples from the first attempt at evaluating a CLD against purposive text data using the specific logic of RIQ analysis. Our knowledge regarding the usefulness of RIQ analysis beyond this context is somewhat limited. For example, one could use RIQ analysis for nonpurposive text data (e.g. newspaper articles, scientific articles, social media posts, etc.). Additional use cases may point to the need for modifications to the RIQ method. Nevertheless, the RIQ method was developed in broader considerations of system dynamics theory and methods, as well as qualitative methods outside of system dynamics, so we expect the RIQ method to be useful more broadly.

For limitations specific to the broader suite of methods and data used in the parent study (Tomoaia-Cotisel, 2018), please see the limitations section of the original work.

Future research

In this article, we have described the RIQ method, we have illustrated its use, and we have discussed its costs and benefits as we experienced them in one case. We find it to be a useful method for tests of structure.

We argue for *enlightened* flexibility in modification and adaptation of this method for specific project's needs and available resources. For example, one can adapt the method to be less rigorous by choosing to omit documenting the interpretation. Or one could adapt the method to be more rigorous by choosing to separately document the interpretation for each link rather than to write one interpretation for the entire causal chain that is expressed. Also, one can choose not to test perception of delays and feedback loops. Or one can choose to test perception of stocks and flows in addition to the elements of structure already called out in this method.

Any changes to the RIQ method will come with their own costs and benefits, and these costs and benefits should be carefully considered before they are implemented. If implemented, we also argue for *documentation of changes* made to the method and post hoc reflection and dissemination of use cases. Additional use cases will prove useful in considering how this promising method can become a best practice.

DOI: 10.1002/sdr

Future research could also explore the operationalization of the RIQ method in different research team settings (e.g. where one is the lone modeler in a team of social scientists, where one is one of several modelers on the team performing the RIQ analysis), for different clients (e.g. ones that are more or less willing to engage in using the RIQ results), and using more feature-rich software tools (e.g. in spreadsheet, database, or modeling software).

This article presents the use of RIQ analysis in late-stage conceptualization, where we evaluated CLD structure against contemporaneous interviews internal to the same organization. Future research could branch out beyond this use; for example, RIQ analysis could be adapted for:

- testing saturation by asking: have we heard everything that is relevant to the problem, or have we ceased to surface conceptualization errors?
- early-stage conceptualization - as a front-end tool to rigorously develop one's initial CLD.
- testing simulation model structure and behavior against text data
- testing the assumption that the research method is appropriate to the study context. For system dynamics, by asking: do study participants see their problem in terms of dynamic complexity and mental models?
- testing external validity by comparing a CLD to data gathered in a totally new context or in the same context at a different time period.
- discussing with interdisciplinary-research team members: in developing a codebook for the CLD, in discussing alternatives for simulation model structure, etc.
- discussing with diverse stakeholders, opening up the opportunity to evaluate the CLD together in an approachable manner. In the analysis stage, this could provide a qualitative alternative to *interrater reliability* where RIQ tables can be used to support a multistakeholder conversation with the purpose of coming to agreement about how quotations support and/or challenge a dynamic hypothesis.

While we have envisioned future developments as potentially supporting research using system dynamics, we believe that, as it provides a logical skeleton for testing assumptions against text data, there are also potential applications in studies using other methods such as agent-based modeling. We have only scratched the surface for how RIQ analysis could be used in research processes to improve understanding and affect change. For example, one could explore the relative effectiveness of different methods for communicating RIQ results with stakeholders (e.g. in RIQ tables, in quotations where RIQ analysis is used to write the interpretation, as part of an online interactive diagramming tool). As demonstrated here, we are confident that the RIQ method has potential applications beyond what is illustrated in this article.

DOI: 10.1002/sdr

Acknowledgements

This project was supported in part by grant R18HS019136 from the Agency for Health Care Research and Quality as part of its program to study the transformation of primary care, Michael K. Magill, M.D., Principal Investigator. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Health Care Research and Quality. The authors would like to thank the three anonymous reviewers whose thoughtful comments helped improve and clarify our manuscript. We would also like to thank William L. Miller for his contributions regarding mixed methods broadly.

Biographies

Andrada Tomoaia-Cotisel is a Policy Researcher at the RAND Corporation and Professor of Policy Analysis at the Pardee RAND Graduate School. She is also an Honorary Research Fellow at the London School of Hygiene and Tropical Medicine (LSHTM). She teaches and mentors Ph.D. students in mixed-methods approaches to system dynamics modeling and systems thinking. She received her Ph.D. in Health Services Research & Policy from LSHTM. She specializes in developing and applying formal methods bringing the strengths of qualitative and quantitative data to improve conceptualization and validation. Her current work explores dynamic complexity in health service delivery, implementation, and outcomes, as well as the influence of context and resulting variation.

Samuel Allen is a Ph.D. Candidate in The Business School at Worcester Polytechnic Institute (WPI). He is also an NSF Research Trainee with WPI's Circular Economy and Data Analytics Engineering Research for Sustainability Center (CEDAR) and a Visiting Ph.D. student with the Logistics Systems Dynamics Group at Cardiff University. His dissertation research is on the dynamics of supply chain strategy and food waste reduction. His teaching areas include operations management and sustainability and ran the WPI TimeWise simulation lab. He consults on the development of formal methods using qualitative data to complement existing system dynamics methods and in the incorporation of these methods into projects in health services and operations management.

Hyunjung Kim, Ph.D., is a Professor and a Thomas Family Endowed Fellow in the Department of Management, California State University, Chico. Her teaching areas include management, strategic decision-making, and system dynamics modeling. She received her Ph.D. in Public Administration from the Rockefeller College of Public Affairs and Policy, University at Albany.

DOI: 10.1002/sdr

She has been working on development of qualitative methods for systematically eliciting and representing stakeholder mental models. Her system dynamics research and consulting work covers various policy and management areas such as data-intensive organizations, public service delivery, higher-education planning, and monetary policymaking.

David F. Andersen is an Emeritus Professor of public administration and information science at the Rockefeller College of Public Affairs and Policy, University at Albany—SUNY. He holds an AB in Mathematics and Social Sciences from Dartmouth College and his Ph.D. in Management from the MIT Sloan School of Management. He has served as Dean of the Graduate School of Public Affairs and Director of the Rockefeller Institute of Public Affairs and Policy at the State University of New York. Prof. Andersen has been awarded the Jay W. Forrester Prize for the best published work in system dynamics. Professor Andersen is a Past President of the System Dynamics Society. His current work includes mentoring and using system dynamics to explain choices people face when grappling with the COVID-19 pandemic (newfadumfarm.org).

Dr. Zaid Chalabi is a mathematical modeler. He is Honorary Associate Professor in Mathematical Modeling at the Institute for Environmental Design and Engineering, Bartlett Faculty of the Built Environment, University College London. He has research interests in modeling the health impacts of the indoor and outdoor environments and climate change. He is also interested in several mathematical methods allied to modeling including uncertainty, risk and decision analyses, extreme value analysis, and optimization. His current research interests are complexity science modeling methods and their applications including agent-based modeling and system dynamics modeling.

References

Andersen DF, Richardson GP. 1997. Scripts for group model building. *System Dynamics Review: The Journal of the System Dynamics Society* **13**(2): 107–129.

Andersen DF, Richardson GP, Vennix JAM. 1997. Group model building: Adding more science to the craft. *System Dynamics Review* **13**: 187–201.

Andersen DL, Luna-Reyes LF, Diker VG, Black L, Rich E, Andersen DF. 2012. The disconfirmatory interview as a strategy for the assessment of system dynamics models. *System Dynamics Review* **28**(3): 255–275.

Armstrong DJ. 2005. Causal mapping: A discussion and demonstration. In *Causal Mapping for Research in Information Technology*, Nakayama VK, Armstrong DJ (eds). Hershey, PA: Idea Group Publishing,

Barlas Y. 1996. Formal aspects of model validity and validation in system dynamics. *System Dynamics Review* **12**(3): 183–210.

DOI: 10.1002/sdr

Becker HS, Geer B, Hughes E, Strauss A. 1961. *Boys in White: Student Culture in Medical School*. Chicago: University of Chicago Press.

Burchill G, Fine CH. 1994. Working Paper: time versus market orientation in product concept development: empirically-based theory generation.

Burchill G, Fine CH. 1997. Time versus market orientation in product concept development: Empirically-based theory generation. *Management Science* **43**(4): 465–478.

Burchill G, Kim DH. 1993. Working paper: Inductive system diagrams: An empirically based theory generation technique, Alfred P. Sloan School of Management

Crabtree BF, Miller WL. 1999. *Doing Qualitative Research*. Sage Publications, Inc: Thousand Oaks, CA.

Creswell JW, Miller DL. 2000. Dererming validity in qualitative Inqury. *Theory Into Practice* **39**(3): 124–130.

Creswell JW. 2014. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 4th. Thousand Oaks, CA: Sage Publications, Inc.

de Gooyert V. 2018. Developing dynamic organizational theories; three system dynamics based research strategies. *Quality & Quantity* **53**(2): 653–666.

Eker S, Zimmermann N. 2016. Using textual data in system dynamics model conceptualization. *Systems* **4**(3): 28.

Farr WW. 2017. The elements and structure of a group model building case study database. Paper presented at the International System Dynamics Conference.

Fetters MD, Curry LA, Creswell JW. 2013. Achieving integration in mixed methods designs—Principles and practices. *Health Services Research* **48**(6pt2): 2134–2156.

Forrester JW. 1961. *Industrial Dynamics*. MITPress: Cambridge, MA.

Forrester JW. 1980. Information sources for modeling the National Economy. *Journal of the American Statistical Association* **75**(371): 555–566.

Guest G, MacQueen KM, Namey EE. 2012a. Data reduction techniques. In *Applied Thematic Analysis*, MacQueen KM (ed). Sage Publications, Inc: Thousand Oaks, CA.

Guest G, MacQueen KM, Namey EE. 2012b. Introduction to applied thematic analysis. In *Applied Thematic Analysis*, MacQueen KM (ed). Sage Publications, Inc.: Thousand Oaks, CA.

Harding DJ, Seefeldt KS. 2013. Mixed methods and causal analysis. In *Handbook of Causal Analysis for Social Research*, Morgan SL (ed). Handbooks of Sociology and Social Research. Dordrecht: Springer.

Hodgkinson GP, Clarkson GP. 2005. What have we learned from almost 30years of research on causal mapping? Methodological lessons and choices for the information systems and information technology communities. In *Causal Mapping for Research In Information Technology*, Nakayama VK, Armstrong DJ (eds). Hershey, PA: Idea Group Publishing.

Hovmand PS. 2014. Group model building and community-based system dynamics process. In *Community Based System Dynamics*, Hovmand PS (ed). New York: Springer.

Hovmand PS, Andersen DF, Rouwette E, Richardson GP, Rux K, Calhoun A. 2012. Group model-building 'scripts' as a collaborative planning tool. *Systems Research and Behavioral Science* **29**(2): 179–193.

Kim H, Andersen DF. 2012. Building confidence in causal maps generated from purposive text data: Mapping transcripts of the Federal Reserve. *System Dynamics Review* **28**(4): 311–328.

DOI: 10.1002/sdr

Kopainsky B, Luna-Reyes LF. 2008. Closing the loop: Promoting synergies with other theory building approaches to improve system dynamics practice. *Systems Research and Behavioral Science: The Official Journal of the International Federation for Systems Research* **25**(4): 471–486.

Lane DC. 1995. The folding star: A comparative reframing and extension of validity concepts in system dynamics. Paper presented at the International Conference of the System Dynamics Society.

Lane DC. 2015. Validity is a matter of confidence—But not just in system dynamics. *Systems Research and Behavioral Science* **32**(4): 450–458.

Luna-Reyes LF, Andersen DL. 2003. Collecting and analyzing qualitative data for system dynamics: Methods and models. *System Dynamics Review: The Journal of the System Dynamics Society* **19**(4): 271–296.

Miller WL, Crabtree BF, Harrison MI, Fennell ML. 2013. Integrating mixed methods in health services and delivery system research. *Health Services Research* **48**(6pt2): 2125–2133.

Nakayama VK, Armstrong DJ. 2005. *Causal Mapping for Research in Information Technology*. Hershey, PA: Idea Group Publishing.

Oral M, Kettani O. 1993. The facets of the modeling and validation process in operations research. *European Journal of Operational Research* **66**(2): 216–234.

Perlow LA, Okhuysen GA, Repenning NP. 2002. The speed trap: Exploring the relationship between decision making and temporal context. *Academy of Management Journal* **45**(5): 931–955.

Reiss J. 2009. Causation in the social sciences: Evidence, inference, and purpose. *Philosophy of the Social Sciences* **39**(1): 20–40.

Richardson GP. 1999. Reflections for the future of system dynamics. *Journal of the Operational Research Society* **50**(4): 440–449.

Richardson GP. Nov. 7, 2019. Email message to author (SDA).

Richardson GP, Andersen DF. 1995. Teamwork in group model building. *System Dynamics Review* **11**(2): 113–137.

Ryan GW, Bernard HR. 2000. Methods of collecting and analyzing empirical materials: Data management and analysis methods. In *Handbook of Qualitative Research*, Denzin NK, Lincoln YS (eds). Thousand Oaks, CA: Sage Publications, Inc., 769–802.

Ryan GW, Bernard HR. 2006. Testing an ethnographic decision tree model on a National Sample: Recycling beverage cans. *Human Organization* **65**(1): 103–114.

Saeed K. 1992. Slicing a complex problem for system dynamics modeling. *System Dynamics Review* **8**(3): 251–261.

Sterman JD. 2000. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Boston, MA: Irwin/McGraw-Hill.

Sterman JD. 2018a. Fine tuning your causal-loop diagrams-part I. https://thesystemsthinker.com/fine-tuning-your-causal-loop-diagrams-part-i/.

Sterman JD. 2018b. Fine tuning your causal-loop diagrams-part II. https://thesystemsthinker.com/fine-tuning-your-causal-loop-diagrams-part-ii/.

Sterman JD. 2018c. System dynamics at sixty: The path forward. *System Dynamics Review* **34**(1–2): 5–47.

Tashakkori A, Teddlie C. 1998. Mixed methodology: Combining qualitative and quantitative approaches. In *Applied Social Research Methods Series v. 46*. Sage Publications, Inc.: Thousand Oaks, CA.

DOI: 10.1002/sdr

Tomoaia-Cotisel A. 2018. The journey toward the patient-centered medical home: A grounded, dynamic theory of primary care transformation. In *Health Services Research and Policy*. London School of Hygiene and Tropical Medicine. https://researchonline.lshtm.ac.uk/id/eprint/4647856/

Turner BL, Kim H, Andersen DF. 2013. Improving coding procedures for purposive text data: Researchable questions for qualitative system dynamics modeling. *System Dynamics Review* **29**(4): 253–263.

Van Maanen J. 1979. The fact of fiction in organizational ethnography. *Administrative Science Quarterly* **24**(4): 539–550.

Vennix JA. 1999. Group model-building: Tackling messy problems. *System Dynamics Review: The Journal of the System Dynamics Society* **15**(4): 379–401.

Vriens D, Achterbergh J. 2006. The social dimension of system dynamics-based modelling. *Systems Research and Behavioral Science* **23**(4): 553–563.

Yin RK. 2009. *Case Study Research: Design and Methods*. Applied Social Research Methods. **5**, 4th, Thousand Oaks, CA: Sage Publications, Inc.

Appendix A

Online Appendix

Figure A1 below shows how the RIQ method, as used in late-stage conceptualization (and as presented in this paper), fits into a system dynamics process. The RIQ method is found at the starred location: "Comparison and Reconciliation of Structure"). Dashed arrows show processes occurring only in initial stages and solid arrows show processes occurring in any stage. Arrow thickness describes the role in a modeling process. Thick-lined arrows are validation processes. Thin-lined arrows represent inputs that are available to these processes. Colors (non-gray) show how the RIQ method as used in late-stage conceptualization (and as presented in this paper) fits into a system dynamics process. Blue arrows indicate processes prior to RIQ analysis. Tan arrows indicate the inputs to the RIQ analysis. Orange text indicates where the RIQ analysis takes place. The purple arrow indicates the product of the RIQ analysis.

Coloring shows how this method fit into the process of the case study presented in this article. The dashed thin-lined blue arrow indicates that a first set of interview transcripts capturing "Perceptions of System Structure" informed the initial "System Conceptualization." The solid thick-lined blue arrows indicate that initial conceptualization led to the formulation of a CLD based on the purposive text in those interviews. RIQ analysis (starred in this diagram) is a method for "Comparison and Reconciliation of Structure." The solid tan arrows indicate the inputs to be used in RIQ analysis: (1) the CLD as a "Representation of Model Structure" (the solid thick-lined tan arrow) and (2) a validation dataset of purposive text interviews capturing "Perceptions of System Structure" (the solid thin-lined tan arrow). The solid thick-

DOI: 10.1002/sdr

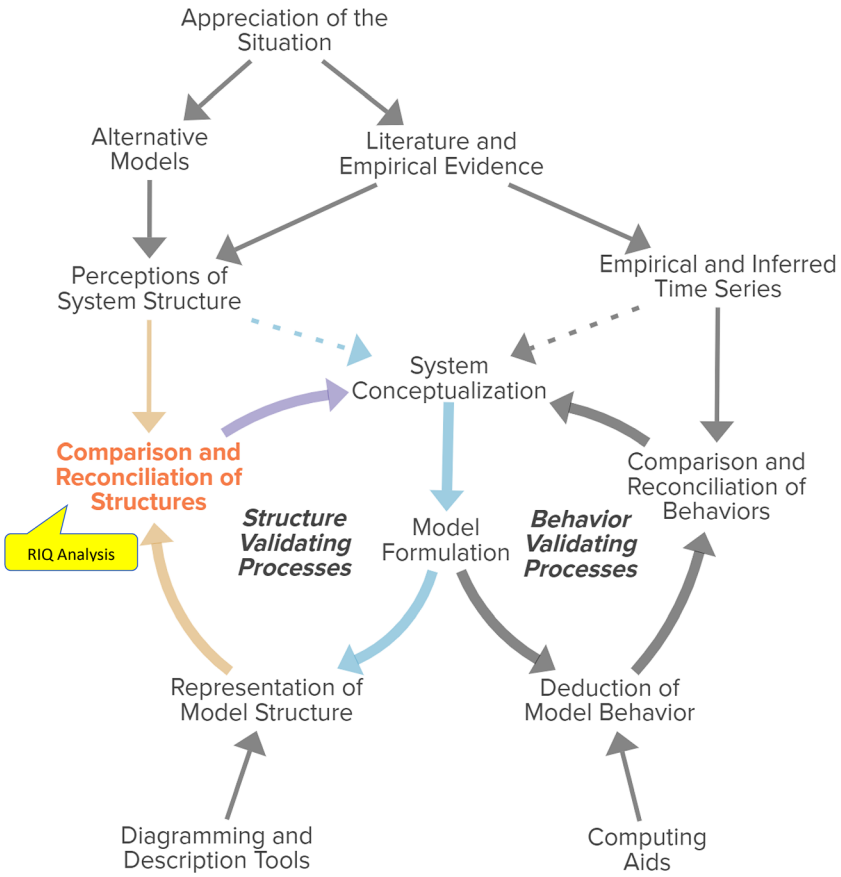

Fig. A1. Location of RIQ Analysis From a System Dynamics Process Perspective. This diagram (adapted from (Saeed, 1992), (Lane, 1995) and (Richardson, 2019)) shows a system dynamics modeling process. Dashed arrows show processes occurring only in initial stages and solid arrows show processes occurring in any stage. Arrow thickness describes the role in a modeling process. Thick-lined arrows are validation processes. Thin-lined arrows represent inputs that are available to these processes. Colors (non-gray) show how the RIQ method, as used in late-stage conceptualization, fits into a system dynamics process. Blue arrows indicate processes prior to RIQ analysis. Tan arrows indicate the inputs to this RIQ analysis. Orange text indicates where this RIQ analysis takes place (also starred in this diagram). The purple arrow indicates the product of this RIQ analysis. An interactive version of this figure is available at: https://kumu.io/sdallen/the-system-dynamics-modeling-process#sd-modeling-process [Color figure can be viewed at wileyonlinelibrary.com]

lined purple arrow indicates that the output of RIQ analysis is an improved "System Conceptualization." In this case, this post-RIQ conceptualization is displayed via a post-RIQ CLD – a CLD that incorporates the changes identified using the RIQ method.

Figure A2 shows the *Trust* loop in Post RIQ versions of the CLD and simulation model. These three diagrams show how this loop was operationalized

DOI: 10.1002/sdr

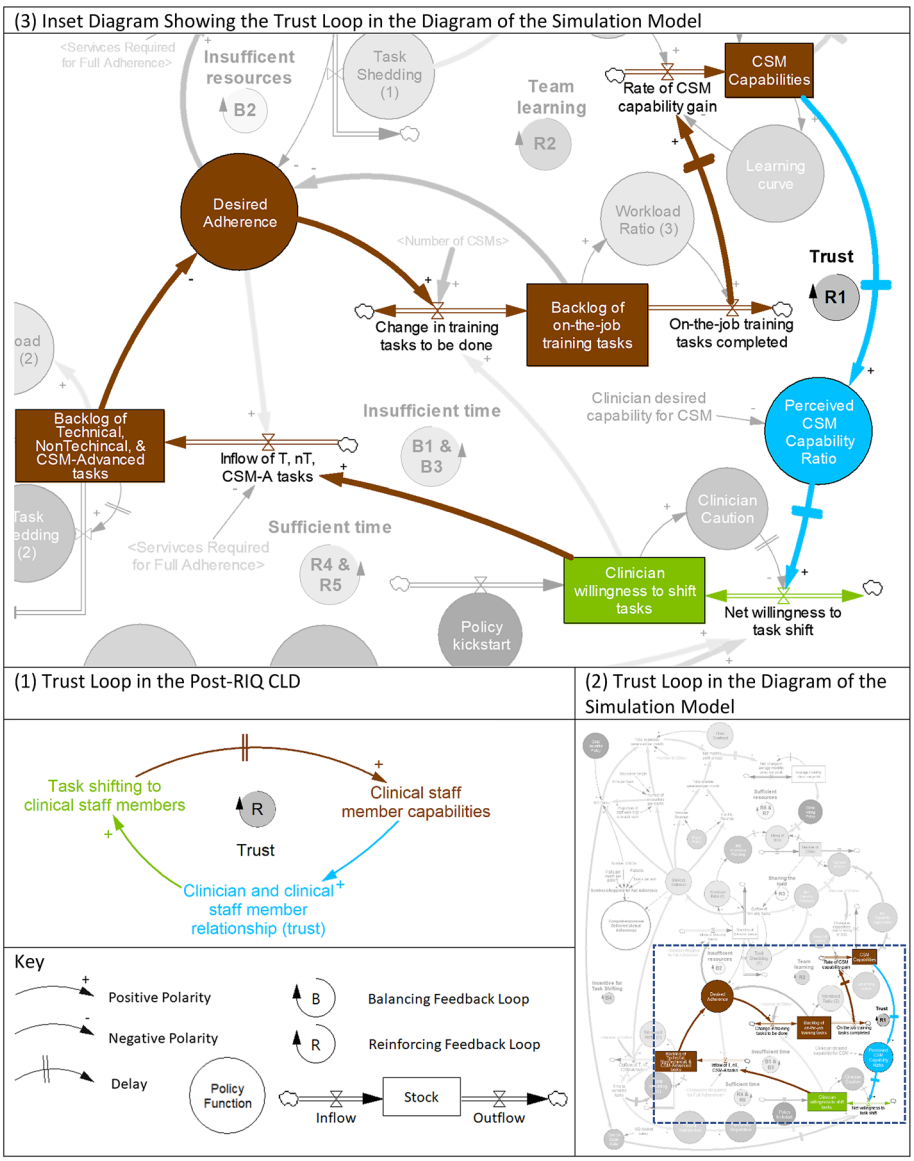

Fig. A2. The Trust Loop as Found in Post-RIQ Diagrams. This figure is adapted from (Tomoaia-Cotisel, 2018) Figure 3.86 with simulation model equations in section 3.4.2. *Trust* loop elements are colored in the Post-RIQ CLD and in the diagram of the simulation model [Color figure can be viewed at wileyonlinelibrary.com]

in the parent study (Tomoaia-Cotisel, 2018). Diagram 1 shows *Trust* loop elements in the Post-RIQ CLD and is colored to help locate its variables and links in Diagrams 2 and 3. Diagram 2 shows the *Trust* loop in context of the simulation model. Diagram 3 provides the operational details of the *Trust* Loop in the simulation model. Below, we describe what happens in this loop when trust increases, using the three colors in Figure A2.

DOI: 10.1002/sdr

- Brown (starting from the green stock in the bottom right): as clinicians become more willing to shift tasks to clinical staff members, their team offers more tasks (i.e. primary-care services) to their patients than before. As more services are consistently provided to patients, the clinicians on the team see that they are improving their adherence to clinical guidelines (i.e. providing more comprehensive care) and this increases their desired adherence. Because this goal requires clinical staff members to have experience with new tasks, it generates new on-the-job training tasks that need to be completed. As teams complete these training tasks, clinical staff members master new capabilities needed to perform the additional tasks shifted to them.
- Blue: Clinicians consider what they perceive to be the capabilities level of the clinical staff members on their team and adjust how much they trust them to perform the shifted tasks.
- Green: Trust in clinical staff members is one of several factors that clinicians consider in adjusting their willingness to shift tasks (i.e. determining how much task-shifting to undertake).

DOI: 10.1002/sdr