Case study of evaluations that go beyond clinical outcomes to assess quality improvement diabetes programmes using the Diabetes Evaluation Framework for Innovative National Evaluations (DEFINE)

Jann Paquette-Warren MSc,[1] Stewart B. Harris CM MD MPH FCFP FACPM,[2] Mariam Naqshbandi Hayward MSc[1] and Jordan W. Tompkins MA[1]

[1]Program Coordinator, [2]Professor, Centre for Studies in Family Medicine, Western Centre for Public Health and Family Medicine, Department of Family Medicine, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada

Abstract

Rationale, aims and objectives Investments in efforts to reduce the burden of diabetes on patients and health care are critical; however, more evaluation is needed to provide evidence that informs and supports future policies and programmes. The newly developed Diabetes Evaluation Framework for Innovative National Evaluations (DEFINE) incorporates the theoretical concepts needed to facilitate the capture of critical information to guide investments, policy and programmatic decision making. The aim of the study is to assess the applicability and value of DEFINE in comprehensive real-world evaluation.

Method Using a critical and positivist approach, this intrinsic and collective case study retrospectively examines two naturalistic evaluations to demonstrate how DEFINE could be used when conducting real-world comprehensive evaluations in health care settings.

Results The variability between the cases and the evaluation designs are described and aligned to the DEFINE goals, steps and sub-steps. The majority of the theoretical steps of DEFINE were exemplified in both cases, although limited for knowledge translation efforts. Application of DEFINE to evaluate diverse programmes that target various chronic diseases is needed to further test the inclusivity and built-in flexibility of DEFINE and its role in encouraging more comprehensive knowledge translation.

Conclusions This case study shows how DEFINE could be used to structure or guide comprehensive evaluations of programmes and initiatives implemented in health care settings and support scale-up of successful innovations. Future use of the framework will continue to strengthen its value in guiding programme evaluation and informing health policy to reduce the burden of diabetes and other chronic diseases.

Introduction

Background

Many efforts to improve health systems were made over the past decade to mitigate the global rising prevalence rates and associated financial and clinical burden of diabetes [1,2]. In Canada, the United States and numerous other countries, national and/or regional diabetes strategies were funded to enhance health promotion, disease prevention and disease management [1,3–5]. In parallel, extensive health care reforms fostered a proactive chronic disease management approach and programmes were implemented to support the re-design of diverse health care organizations with more emphasis on team-based care, electronic infrastructure and care delivery incentives [6,7]. However, evaluations of these efforts have, in general, produced insufficient evidence to inform future investments and health system improvements [8–11]. When evidence of successful innovations was available, there were many barriers to participating in knowledge translation to scale-up successes [4,12,13]. In 2008, Borgermans *et al.* suggested that the shortfall of evidence to guide investments and knowledge translation in part exists because a comprehensive and systematic diabetes evaluation framework was not available [14].

Evaluation frameworks

In 2014, Paquette-Warren *et al.* examined existing health-related and diabetes-specific evaluation frameworks and found a continued need for a framework that would provide step-by-step guidance from study conceptualization, to consideration of current methodological options, knowledge exchange and translation [10]. The authors stated that the main overall limitations of existing frameworks was the lack of clear processes to: (1) capture quality and meaningful data using multi-level indicators (e.g. key diabetes indicators for surveillance, indicators related to the system or environment including structural and organizational features, and indicators for chronic care and quality of care related to different aspects of health); and (2) explore the causal relationships between investments, programmes and process/outcomes [10]. Other key aspects suggested as crucial were pulling together the strengths of existing frameworks such as lists of indicators from different sectors for comprehensiveness and finding balance among methodological rigour, cost and feasibility.

The Diabetes Evaluation Framework for Innovative National Evaluations (DEFINE) was developed as an attempt to overcome limitations of current frameworks [15,16]. DEFINE has yet to be tested to determine its strengths and limitations, but it incorporates key factors such as acquisition of comprehensive evidence necessary to critically inform and enable decision makers, as well as active participation in knowledge translation and exchange for scalability of positive health system change [9,10,14,17]. Unlike previous evaluation frameworks, DEFINE includes a stepwise evaluation approach with suggested processes to facilitate assessment beyond traditional system performance by exploring the relationships between programmes and diabetes outcomes. It includes a comprehensive list of multi-level indicators related to the Organization of Healthcare, Healthcare Delivery, Environment and the Patient that are inclusive of, but not limited to, clinical processes and outcomes. Furthermore, it stresses the use of mixed-methods and integrated knowledge translation through the involvement of stakeholders in setting evaluation goals and prioritization of indicators to accommodate real-world evaluation constraints such as timelines, data availability and financial and human resources. A more detailed description of existing frameworks and the development of DEFINE is published [10] and the current DEFINE framework and associated interactive tools are available at http://tndms.ca/research/define/index.html.

Aim

It is hypothesized that using DEFINE's prescribed steps and guiding principles could yield more comprehensive and meaningful evaluation results to better inform decision makers regarding fund allocation, health care policy and health system change aimed at decreasing the burden of diabetes. The purpose of this case study was to examine how DEFINE could be applied as an evaluation tool by studying two previously conducted and reported naturalistic comprehensive evaluations [the evaluations of the Partnerships for Health (PFH) Program and the Quality Improvement and Innovation Partnership (QIIP) Program]. The authors retrospectively explored the similarities and differences among the processes, designs and methods undertaken in these cases using a framework analysis approach to expand the knowledge of DEFINE's usability, strengths and limitations and to provide an example of how DEFINE could be used to conduct future real-world comprehensive evaluations in health care settings [18,19]. A critical and positivist approach was used for this intrinsic and collective case study.

Case descriptions

Case 1

PFH was a chronic disease prevention and management framework demonstration project with a focus on quality improvement [20–22]. Implemented between 2008 and 2011, the programme targeted inter-professional primary health care teams (family doctors, practice-based allied providers and administrative staff and community-based allied providers and administrative staff) and used diabetes as a proxy to improve team-based chronic care in the region of South western Ontario, Canada.

Case 2

QIIP [23,24] was implemented from 2008 to 2010 and targeted inter-professional primary health care teams across the entire province of Ontario, Canada. The overall goal of this quality improvement programme was to develop a high-performing primary health care system focused on chronic disease management, disease prevention and improved access to care using diabetes management, colorectal cancer screening and patient access as proxies.

Both programmes were founded on the principles of the Expanded Chronic Care Model [25,26] and utilized the Institute for Healthcare Improvement Breakthrough Series adult learning model [27]. Through a series of learning sessions, both programmes brought health care teams together to learn about their target areas (e.g. diabetes) and associated clinical practice guidelines, as well as mechanisms to improve chronic care using a team-based approach. Between the learning sessions (action periods of approximately 3 months), teams returned to their practices to test change ideas on a small scale prior to large-scale implementation. Albeit to varying degrees, the programmes supported the work of participants with quality improvement coaches, teleconferences with programme leaders and invited expert speakers, website forum communications, the sharing and uploading of improvement data and information technology (IT) training. Each programme highlighted the importance of scaling-up their successes by spreading lessons learned to colleagues in health care settings.

The main differences between the programmes were that: (1) QIIP was more extensively implemented as a provincial initiative whereas PFH targeted a sub-provincial region; (2) PFH focused primarily on diabetes, process mapping and optimizing electronic population-based surveillance; whereas, QIIP focused on diabetes, colorectal cancer screening and office access and efficiency; (3) PFH required the development of partnerships between primary health care team members and external diabetes, mental health and other community care providers; (4) PFH offered more continuous and extensive onsite and offsite IT support; and (5) the PFH

evaluation was conducted prospectively to guide programme implementation and provide feedback on critical changes needed in real time.

In both cases, programmatic impact was studied using comprehensive external mixed-methods evaluations. The variability in the designs used for these evaluations are described in detail as part of the results of this article. Details about the evaluation results of the PFH and QIIP evaluations can be found in the published literature [20,22–24].

Methods

Based on Crowe *et al.* [19], the design for this case study was: (1) intrinsic – focused on learning more about the implementation processes used to conduct comprehensive evaluations of quality improvement diabetes programmes; and (2) collective – used more than one case to broaden the understanding of the complexity of doing real-world evaluation research to inform investments into health system change. A critical approach was used to question assumptions about the cases using an evaluation approach or processes that produced comprehensive and meaningful results for decision makers. To maximise learning from factors unrelated to the assumptions, a positivist approach was used to explore pre-identified variables in the findings and to test the theoretical steps and guiding principles embedded in DEFINE.

Case selection and data collection

The cases were selected because they were available in the published literature and unique in their comprehensive approach to evaluating chronic disease quality improvement programmes. Access to the details about these evaluations, beyond those provided in the published literature such as interim and final reports or raw data or implementation notes and documentation or case investigators, were available to the authors as research members of the Centre for Studies in Family Medicine at Western University. The cases evaluated programmes that were very similar in terms of their goals, structure and implementation, yet the evaluation design and implementation used to assess their impact revealed some interesting similarities and differences. This made these cases ideal for studying the evaluation process in relation to the theoretical step-wise approach of DEFINE.

Data analysis

A framework approach [19] was used starting with familiarization of the cases' documents, followed by identifying a thematic framework to analyse the content [i.e. selection of DEFINE [10]], indexing and charting the content of individual materials to the framework steps and sub-steps, and then mapping and documenting interpretations as part of the triangulation process within and between the cases [28]. The trustworthiness of the findings was enhanced through the use of a theoretical framework (alignment of the data to the framework steps and sub-steps), by validating the findings with evaluation investigators (iterative individual and team review and analyses) and transparency related to case selection, researcher involvement and explicit description of interpretations/conclusions [18,26,29].

Results

DEFINE – Goal A: guide comprehensive evaluation

The first goal of DEFINE is to guide comprehensive evaluation (Fig. 1). Quality improvement programmes designed to improve health care delivery can be complex, with wide-ranging stakeholder groups, target populations, resources and expected outcomes. This can make it challenging to conduct a meaningful evaluation. Under Goal A, Step 1 of DEFINE is focused on ensuring that the right people are involved to formulate the right questions and approach to obtain valuable information about the impact of a programme. To accomplish this, five sub-steps outline important action items: exploring the possible interactions and causal linkages among the programme and the determinants of health (1.1), identifying implementation processes and intended outcomes in a logic model (1.2), assessing the value of the evidence being sought and discussing indicators of interest (1.3), weighing the strengths and limitations of study designs (1.4), and selecting evaluation questions and an approach that is feasible and will best achieve the goals of stakeholders (1.5) [16].

Step 1: Work iteratively with stakeholders to identify evaluation goals

Both the PFH and QIIP evaluations brought multiple stakeholders together to set evaluation priorities and to provide high-level oversight of the evaluation design, implementation and interpretation of results. Ad hoc and regular meetings were held with stakeholders including government representatives, funders, external evaluation team members, programme administrators and implementers. QIIP also included representatives of participating doctors and primary health care team administrators. PFH did not include programme participants because the evaluation was done concurrently with programme implementation, but community health care representatives and clinical experts were included.

Each evaluation team developed a programme logic model to ensure a common understanding of the programme and to identify potentially meaningful evaluation results. In both cases, an iterative process was used to summarize programme activities and anticipated outcomes and to explore the causal linkages and interactions between them. During this process, open communication and collaboration between the external evaluation team and the stakeholders were necessary to confirm an accurate summary of the programmes, and served as the building blocks to establish positive relationships prior to implementing the evaluation. Main evaluation goals were set in both cases and specificity and feasibility was considered when exploring sub-objectives and potential methodologies. Objectives deemed by the stakeholders to not be feasible were included in the logic models to help gain a better understanding of the targeted and/or anticipated intermediate and long-term outcomes of the programme, and to ensure that short-term outcomes and indicators would be explored in order to make inferences about potential programmatic impact over time.

In PFH, the main goal set by the stakeholders was to determine the programmatic impact on patients living with diabetes by examining whether implementation and participation in PFH: (1) resulted in a change in chronic care delivery; and (2) improved diabetes care processes and outcomes. Emphasis was placed on gathering
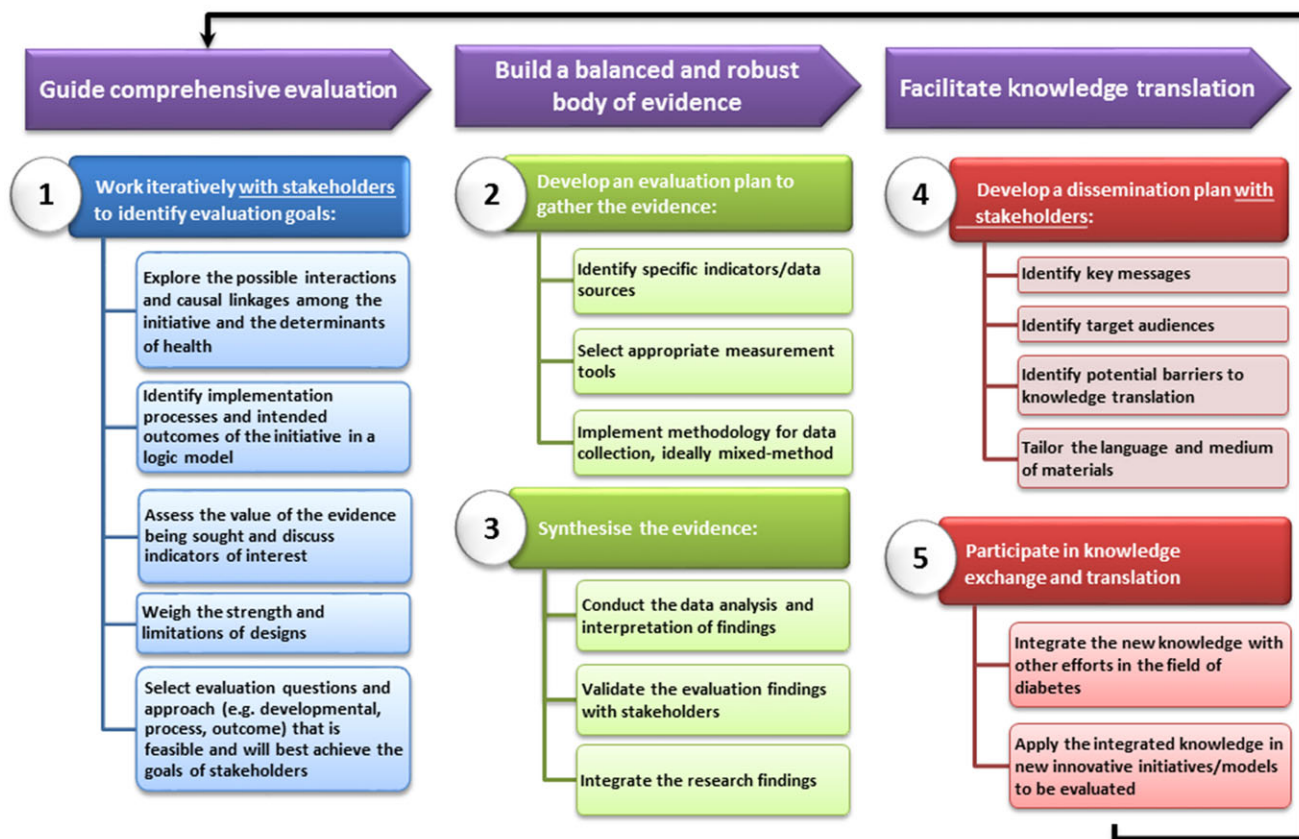
Figure 1 Diabetes Evaluation Framework for Innovative National Evaluations 5-Step Evaluation Framework.

data for each component of the chronic disease prevention and management framework [21]: approach and commitment to chronic care, partnerships and coordination among primary and community health care sites and the patients, the role of patients in care, care delivery and information technology systems and procedures and changes in process and clinical outcomes.

QIIP's main evaluation goal was to assess the impact of the programme on the development of a high-performing primary health care system focused on chronic disease prevention and management, and improved access to care. It centred on three core programme topics: diabetes management, colorectal cancer screening and office access and efficiency. An assessment of programme implementation was included as well as a focus on measurement of the application of teachings to other clinical situations, the relationship between team and practice characteristics and clinical outcomes, and clinical changes over time.

The main difference between the cases was that PFH concentrated on measuring changes occurring within programme participants (pre-post) whereas QIIP compared participant results to a control population. This led to the selection of different designs, methodologies and data sources for gathering evidence.

DEFINE – Goal B: build a balanced and robust body of evidence

The literature calls for more robust evidence related to the description and impact of programmes [8,30] which is DEFINE's second goal: to build a balanced and robust body of evidence (Fig. 1). Multiple sequential steps and sub-steps are housed under Goal B: Step 2, developing an evaluation plan to gather the evidence. This is accomplished by identifying specific indicators and data sources (2.1), selecting appropriate measurement tools (2.2), and ideally implementing a mixed-methods methodology to collect the data (2.3). Step 3 follows with the synthesis of the evidence through data analysis and interpretation of findings (3.1), validation of findings with stakeholders (3.2), and integration/triangulation of the findings (3.3) [16].

Step 2: Develop an evaluation plan to gather the evidence
PFH and QIIP evaluators both used a logic model to guide the identification of specific multi-level indicators and data sources (e.g. patients, providers, programme administrators, patient charts, surveillance system, etc.) and to develop a comprehensive mixed-methods design. This was done by aligning indicators, data

sources and methodologies to programme activities and anticipated outcomes listed in the logic model. To ensure comprehensiveness, evidence was sought to demonstrate impact and to provide information about how this impact was made possible by including both process and outcomes evaluation components (i.e. linking the programme activities to outcomes). The design, sampling and participant eligibility for each method were unique to each evaluation because: (1) the PFH evaluation was conducted prospectively, while the QIIP evaluation was conducted retrospectively; (2) QIIP had a smaller evaluation timeline and budget, as well as a larger target population and geographical region; and (3) both had different stakeholders with specific goals and interests. Each design presented advantages and limitations to be mitigated to enhance the rigour, credibility and trustworthiness of the findings.

PFH evaluation activities were selected to capture the complexity and variability within the programme structure, implementation process, programme participants (i.e. process evaluation) and the programme's impact on the care teams' knowledge and approach to care and clinical processes and outcomes (i.e. outcome evaluation). Conducting the evaluation prospectively with implementation observation and parallel data collection strengthened the overall design and provided an opportunity to support programme fidelity, as well as to identify necessary programme changes. A pre–post matched design was selected for the chart audit of patients with type 2 diabetes to collect clinical process and outcomes data. The surveys also used a pre–post matched design to capture the perspective of programme participants and patients on diabetes care processes, care coordination, participation and outcomes. Post-data were limited to the programme timeline and funding; therefore, rigour was enhanced by seeking similar data from multiple sources (i.e. patient chart information, practice-based administrators and providers, community-based administrators and providers, and patients living with diabetes). Pre–post data were supplemented with post-only in-depth interviews and focus groups data (e.g. perspective of programme implementers, participants and patients about their experience in caring for people with diabetes or living with diabetes and obtaining care for themselves, team-functioning, key lessons learned in the programme, spread and sustainability and self-management). Lastly, the qualitative data augmented the participant observation and programme documentation (log books, hand-outs and reports) process evaluation data. Using multiple methods and sources enabled the evaluators to capitalize on the strengths of each method, and gain a broader understanding of the impact of the programme.

The QIIP evaluation was conducted retrospectively using a multi-measure, mixed-methods, pre–post controlled design to capture the complexity and breadth of this real-world programme. The methodological rigour of the design was increased by including a matched-control for the chart audit and advanced access survey components of the evaluation. Post-only surveys (with programme participants) and interviews (with programme implementers and participants) were used to capture data related to short-term outcomes listed within the logic model such as team functioning, capacity to implement quality improvement strategies and interactions among programme participants and coaches. In addition, a process evaluation was done to assess whether the programme was implemented as intended, identify changes made to better meet the needs of the participants and whether it was deemed effective as a

programme. Comparisons were made between the original programme plans and implementation process through analysis of interview and survey results as well as programme documentation (e.g. meeting minutes, attendance sheets and learning session agendas). After discussions between the stakeholders and the evaluation team, the evaluation plan was expanded to include a health administrative data analysis component (controlled pre–post study design). This further strengthened the overall design by creating an opportunity to compare the billing data of consenting QIIP doctors to all other doctors practising in the same funding model in the province.

Regardless of the final designs used in these evaluation cases, researchers were dedicated to capturing evidence that met the evaluation goals set by the stakeholders and focused on comprehensiveness by identifying multiple mixed-methods process and outcome indicators. Table 1 shows that indicators were included across the multiple levels and priority indicators identified in the DEFINE determinants of health schematic and priority indicator list. Also important was the emphasis placed on feasibility with the use of best practice strategies to enhance rigour in the designs. In these cases, similar strategies were used: (1) the use of an external evaluation team; (2) assessment of programme impact across multi-levels of indicators aligned to programme activities, short-term outcomes, data sources and measurement tools in a logic model; and (3) a comprehensive process and outcomes evaluation using a mixed-methods approach.

Step 3: Synthesise the evidence

The complexity of the comprehensive, multi-method/multi-indicator approach used in PFH and QIIP necessitated extensive iterative and multi-step data analysis and synthesis. Data analyses for all methods were first completed independently. Then integration of the results was done using a convergence triangulation approach [28] to build on the strengths of each method and to add rigour to the evidence. In PFH, convergence triangulation was facilitated by aligning the results to the anticipated outcomes listed in the logic model and by displaying the results of each method side-by-side. This activity provided insight into key factors and variations within the results, and allowed the evaluators to make inferences about the programmatic impact. In QIIP, the logic model outcome statements were grouped into key sections such as capacity building, clinical process and outcomes, and team functioning and aligned with the data available from each method. Summary statements for each key section of the logic model were included to highlight the main findings of each methodological component, as well as triangulation statements to denote when qualitative and quantitative findings corroborated and/or converged. This provided a more complete picture of the results and allowed the evaluation team to draw from all findings to inform interpretation and conclusions about the programme.

In both evaluations, preliminary findings were presented to the stakeholders before any final conclusions were drawn about the programmes. Where results did not match anecdotal information or the perspective of stakeholders, the data were revisited to confirm accuracy. This validation process was a valuable exercise to ensure accuracy of the results and to reassure stakeholders that they were verified. This led to slight changes in the reports and additional analyses, though it did not change the main findings of the evaluations.

Table 1 Multi-level indicators included in PFH and QIIP categorized by the DEFINE levels and priority indicators
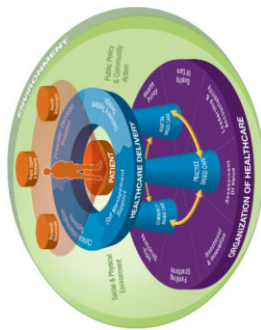
The DEFINE Determinants of Health Schematic displays the relationship among medical and non-medical determinants of health categorized into four levels: Patient, Healthcare Delivery, Organization of Healthcare, and Environment. Priority indicators have been identified for each level. Indicators included in the PFH and QIIP evaluations are marked below. Specific indicators, measurement tools, and methods varied between the evaluations. Additional indicators were included in both evaluations to gather evidence regarding knowledge and application of quality improvement and care redesign methodologies, opportunities and actions, as well as program involvement and perceived value.

Define priority indicators

Patient level		
PFH	**QIIP**	**Health Status & Well-being**
✓	✓	• Glycemic control
✓	✓	• Hypoglycemia
✓	✓	• Anthropometric measures
✓	✓	• Cholesterol/lipid profile
✓	✓	• Medication usage
✓	✓	• Organ specific dysfunction/disease
✓		• Mental health
✓		• Functional capacity
✓		• Quality of life
✓		• Patient experience
PFH	**QIIP**	**Health behaviours**
✓		• Self-efficacy
✓		• Self-care
PFH	**QIIP**	**Personal resources**
✓		• Education/literacy/enablement
✓		• Social support/caregivers
✓		• Employment/socio-economic/health insurance
PFH	**QIIP**	**Organization of healthcare level**
✓		• Effective organizational and funding structure(s)
✓		• Health service utilization patterns
✓		• Accessible health services
✓		• Accountability related to quality of care and system performance

Healthcare delivery level		
PFH	**QIIP**	• Self-management support
✓	✓	• Patient-centred self-management action plans
✓	✓	• Health promotion and education
PFH	**QIIP**	• Delivery system design
✓	✓	• Effective and safe systems or structures
✓		• Focus on patient-centred care and patient interactions
✓		• Alignment of services and continuity of care
PFH	**QIIP**	• Provider decision support
✓		• Routine feedback reports
✓	✓	• Evidence-based guidelines embedded into daily practice
PFH	**QIIP**	• Clinical information systems
✓	✓	• Systematic tracking of clinic data
✓	✓	• Effective use of multi-functional electronic/ medical/health records
✓		• Access to health information

Environment level		
PFH	**QIIP**	• Social environment
✓		• Political/leadership commitment to establishing healthy and equitable socioeconomic, living and working conditions particularly for people with low socio-economic status, aboriginal peoples, immigrants, etc.
PFH	**QIIP**	• Community action
✓		• Existence of accessible system or structures (e.g. Housing, transportation, justice, employment) that are stable, secure, safe, and patient-centered (i.e. Stimulating, satisfying, and enjoyable)

Note: PFH = Partnerships for Health Program; QIIP = Quality Improvement and Innovation Partnership Program.

DEFINE – Goal C: facilitate knowledge translation

Integrated knowledge translation is an integral and essential component of the entire DEFINE framework such as working with stakeholders to set evaluation goals and to verify the accuracy of evaluation results under Goals A and B; however, special emphasis is placed on programme-end knowledge translation in Goal C to ensure that evaluation findings are translated into action (Fig. 1). Without successful knowledge translation, the lessons learned, gaps and successes identified through evaluation cannot effectively be disseminated to knowledge users, funders and the public, which prevents opportunities to inform future programme planning or evaluation. DEFINE encourages two steps to facilitate programme-end knowledge translation: Step 4 – the development of a dissemination plan with stakeholders [select key messages (4.1), identifying potential target audiences (4.2), anticipating and mitigating potential barriers to knowledge translation (4.3) and tailoring the language and medium of the materials for the audience (4.4)], and Step 5 – participating in knowledge exchange and translation to share new information and integrate new knowledge into the field of diabetes (5.1) and planning new innovative programmes and/or evaluations (5.2) [16].

Step 4: Develop a dissemination plan with stakeholders
Both PFH and QIIP were limited in their documentation of programme-end knowledge translation efforts. There is some indication in meeting minutes that potential target audiences, barriers and mediums were considered, but no concrete evidence is available to describe brainstorming and decisions made regarding a dissemination plan.

Step 5: Participate in knowledge exchange and translation
Internal and external strategies were used to optimize project-specific knowledge exchange and continued collaboration or information sharing between the evaluation teams and the programme stakeholders. Internal dissemination activities included reports to stakeholders and reports or presentations to participants. External dissemination strategies in both cases included presentations to government and local health agencies, 10 to 12 professional presentations or posters each, and as of September 2015, two PFH and two QIIP scholarly publications [20,22–24]. In PFH, presentations to the broader health care community in the region and province also occurred. It was unclear from the case information available to the authors how much knowledge exchange occurred independent of the evaluation teams.

From an evaluation perspective, knowledge translation was accomplished through the integration of the lessons learned in PFH into the design of the QIIP evaluation. Subsequently, both cases contributed to the development of FORGE AHEAD, a community-driven, culturally relevant quality improvement research and evaluation programme in Indigenous communities across Canada (http://tndms.ca/forgeahead/index.html). Also, the PFH and QIIP evaluations contributed to the iterative development of DEFINE by systematically integrating the hands-on experience of the PFH and QIIP evaluations with broader evaluation research, methodological and theoretical literature. This reinforced the importance of a comprehensive, mixed-methods evaluation approach capable of informing programme and policy development with the goal of improving the care and health of people living with diabetes. From a programme perspective, there is evi- dence of sustained efforts, but limited information about knowledge to action or translation to other setting or contexts.

Discussion

Investment in health care reform to support better alignment of health care services and delivery to the needs of patients living with chronic diseases continues to be a priority given the ongoing rise in the prevalence, and clinical and financial burden of chronic diseases on the health care system and society [1,2]. However, without more comprehensive evaluations, there is a lack of evidence to inform and support investment into strategies that are successful at improving health care delivery for patients with chronic illnesses [18,26,31]. DEFINE was developed to guide comprehensive evaluation of diabetes prevention and management strategies, and to facilitate policy innovations that could improve diabetes care. The framework builds on the strengths of existing frameworks and substantiates it with the real-world experience of conducting comprehensive evaluations [10]. Specifically, it includes clear steps to facilitate the identification of evaluation goals, selection of meaningful multi-level indicators, gathering of quality data using mixed-methods and a mechanism to explore the causal relationships between investments, programmes and outcomes, which were highlighted as key drawbacks of pre-existing frameworks [10].This study used DEFINE to examine two cases that assessed the impact of quality improvement diabetes programmes and provides insights into how this theoretical framework can be applied as an evaluation tool to shape future real-world evaluations.

To learn more about the potential value of DEFINE's prescribed steps and guiding principles in gathering better evidence to inform decision makers, the published articles and other documentation from the PFH and QIIP evaluations were studied retrospectively and information was aligned to the three main goals and guiding steps of the framework. This exploration of the process, design and methods used in each case revealed under Goal A of DEFINE that both PFH and QIIP cases demonstrated effort in bringing multiple and diverse stakeholders together to work iteratively to summarise programme details, investigate potential interactions and causal linkages among programme elements and anticipated outcomes, and to identify meaningful and feasible evaluation goals and objectives (were of integrated knowledge translation). Of special interest are the relationships developed between the evaluation team and the stakeholders during this process and its perceived beneficial influence during evaluation implementation, completion and end-programme knowledge translation phases. This confirms previous suggestions that collaborative work of this nature may serve as a judicious platform to conduct a successful evaluation [29,32]. Prospective use of the DEFINE tools such as the worksheets related to stakeholder engagement, programme description and logic model development is needed to determine the strengths and limitations of this aspect of the framework and further assess the benefit of these processes in conducting evaluations.

For Goal B of DEFINE, the comprehensiveness of the P4H and QIIP evaluations included key action items associated with building a balanced and robust body of evidence. Although different between the cases, both identified multi-level priority indicators across the four levels of the DEFINE Determinants of Health Schematic included multiple data sources (e.g. stakeholders, pro-

viders, interviews, surveys, charts), and implemented mixed-methods designs to collect data. This aligns well with suggestions that researchers should use sophisticated mixed-methods designs to study interventions and their relationships to lived experiences and impact because no ideal design exists [9]. Furthermore, synthesis of the evidence through data analysis and interpretation of findings, validation of the evaluation findings with stakeholders and integration of the research findings using triangulation occurred in both cases. It was evident that steps 2 and 3 of DEFINE can be accomplished to help gather data that can answer the research question(s) and goals outlined by stakeholders under Goal A. Consistent with reports by Hayes *et al.* [32], the use of the logic model appeared to be key in enabling the alignment of goals, research questions, indicators and evidence. Also, the use of mixed-methods and multiple data sources strengthen the credibility and trustworthiness of the findings and provided a comprehensive understanding of the programmes and their impact [19]. It remains to be seen if having access to the DEFINE tools such as the All-inclusive and Priority Multi-level indicator sets and the DEFINE Determinants of Health Schematic could facilitate the exploration of possible causal linkages and the selection of indicators and tools that would enhance the comprehensiveness of the evaluation and evidence gathered.

Finally for Goal C of DEFINE, the findings confirm the importance and presence of integrated knowledge translation throughout the PFH and QIIP evaluations, not only to develop and implement the evaluation plan, but also to interpret and validate the results. The opportunity for stakeholders to share their views and interpretation of the findings followed by the steps taken by evaluators to re-examine the data with stakeholders' perspectives in mind instilled confidence in the accuracy of the results. However, it is unclear to what extent stakeholders were involved in the development of an appropriate dissemination plan (Step 4) and integration of knowledge gained in new and innovative programmes (Step 5). That said, these cases exemplify the concepts of applying core competencies to advancing the field of evaluation [29]. There exists many barriers, such as institutional, regulatory and financial to overcome in order to encourage more widespread learning and scale-up of successful innovations [4]. Given the limited evidence in both cases to support the involvement of all the stakeholders in a dedicated process to support knowledge exchange and translation, one could suggest that the inclusion of programme-end knowledge translation as a main goal in DEFINE with associated tools and worksheets may help evaluators to examine more closely their role in knowledge translation, and the stakeholders needed to effectively implement lessons learned in new contexts. The guiding principles and processes related to integrated and end-of-programme knowledge translation throughout the framework may be the most critical new addition to evaluation frameworks in realizing the goal of producing better evidence to support superior decision making to address the burden of diabetes.

Limitations

Because PFH and QIIP were large and comprehensive programmes focused on disease management, this case study provided insight into the applicability of the guiding steps and underlying principle of DEFINE. However, the inclusive nature and built-in flexibility of DEFINE, which accommodates its application irrespective of the size, type or goals (health promotion, disease prevention or management) of diabetes programmes, could not be assessed by this intrinsic and collective case study. Doing an instrumental collective case study [19] by applying DEFINE to a variety of programmes should be conducted to test the applicability and generalizability of DEFINE. That said, certain aspects of DEFINE's built-in flexibility were apparent, like how it could be used to conduct prospective and retrospective studies, as well as support the use of different evaluation designs, research methods and indicators in a comprehensive way to meet the goals of diverse stakeholders. Given that comprehensive evaluations such as PFH and QIIP are rare, using more commonly reported evaluation approaches may further support the need for and value of DEFINE [9,11,26,31]. Using P4H and QIIP demonstrated that the comprehensiveness of the framework is feasible in real-world research. Finally, DEFINE could be adapted and expanded to guide the evaluation of a variety of chronic disease programmes. Future work to expand the framework and test it on initiatives that target other chronic diseases is needed.

Conclusion

Retrospective examination of PFH and QIIP demonstrated how DEFINE could be used to structure future comprehensive evaluation of programmes implemented in health care settings. DEFINE's stepwise approach to capturing accurate and meaningful data using multi-level indicators and exploring the causal relationships between investments, programmes and outcomes was exemplified in this case study. More work is needed to assess the applicability of DEFINE on programmes of different types and sizes, and to explore its adaptability to programmes targeting various aspects of chronic care including health promotion, disease prevention and management, as well as other chronic diseases. Also, a prospective study of the strengths and limitations of the DEFINE tools and worksheets will demonstrate more clearly the benefits of the framework in improving evaluation efforts. As previously reported [10], the framework will continue to be refined and strengthened as it gets applied to other diabetes programmes and chronic diseases, and as the body of research on its utility and applicability grows. DEFINE has the potential to become a standardized tool in guiding programme evaluation and informing health policy to reduce the burden of diabetes and other chronic diseases.

Acknowledgements

References

1. Canadian Diabetes Association. Diabetes: Canada at the Tipping Point [Internet]. Canadian Diabetes Association; 2011 pp. 1–59. Available at: http://www.diabetes.ca/documents/get-involved/WEB_Eng.CDA _Report_.pdf (last accessed 17 September 2015).

2. World Health Organization. 10 facts about diabetes [Internet]. 2014. Available at: http://www.who.int/features/factfiles/diabetes/en/ (last accessed 17 September 2015).

3. Glazier, R. H., Kopp, A., Schultz, S. E., Kiran, T. & Henry, D. A. (2012) All the right intentions but few of the desired results: lessons on access to primary care from Ontario's patient enrolment models. *Healthcare Quarterly (Toronto, Ont.)*, 15 (3), 17–21.

4. Thoumi, A., Udayakumar, K., Drobnick, E., Taylor, A. & McClellan, M. (2015) Innovations in diabetes care around the world: case studies of care transformation through accountable care reforms. *Health Affairs (Project Hope)*, 34 (9), 1489–1497.

5. OECD. Health at a Glance 2015: OECD Indicators [Internet]. 2015 p. 1–200. Available at: http://www.oecd-ilibrary.org/docserver/ download/8115071e.pdf?expires=1450365145&id=id&accname= guest&checksum=D23D3227E9FF0325E8255199930B5004 (last accessed 16 December 2015).

6. Hutchison, B., Levesque, J. F., Strumpf, E. & Coyle, N. (2011) Primary health care in Canada: systems in motion. *The Milbank Quarterly*, 89 (2), 256–288.

7. McClellan, M., Kent, J., Beales, S. J., *et al*. (2014) Accountable care around the world: a framework to guide reform strategies. *Health Affairs (Project Hope)*, 33 (9), 1507–1515.

8. Chaudoir, S. R., Dugan, A. G. & Barr, C. H. I. (2013) Measuring factors affecting implementation of health innovations: a systematic review of structural, organizational, provider, patient, and innovation level measures. *Implementation Science*, 8 (1), 1–20.

9. Dubois, N., Lloyd, S., Houle, J., Mercier, C., Brousselle, A. & Rey, L. (2011) Discussion: practice-based evaluation as a response to address intervention complexity. *Canadian Journal of Program Evaluation*, 26 (3), 105–113.

10. Paquette-Warren, J., Naqshbandi Hayward, M., Tompkins, J. & Harris, S. (2014) Time to evaluate diabetes and guide health research and policy innovation: the Diabetes Evaluation Framework (DEFINE). *Canadian Journal of Program Evaluation*, 29 (2), 1–20.

11. Schouten, L. M., Hulscher, M. E., Everdingen, J. J., van Huijsman, R. & Grol, R. P. (2008) Evidence for the impact of quality improvement collaboratives: systematic review. *BMJ (Clinical Research Ed.)*, 336 (7659), 1491–1494.

12. Grimshaw, J. M., Eccles, M. P., Lavis, J. N., Hill, S. J. & Squires, J. E. (2012) Knowledge translation of research findings. *Implementation Science*, 7 (1), 50.

13. Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I. & Petticrew, M. (2013) Developing and evaluating complex interventions: the new Medical Research Council guidance. *International Journal of Nursing Studies*, 50 (5), 587–592.

14. Borgermans, L. A., Goderis, G., Ouwens, M., Wens, J., Heyrman, J. & Grol, R. P. (2008) Diversity in diabetes care programmes and views on high quality diabetes care: are we in need of a standardized framework? *International Journal of Integrated Care*, 8 (Journal Article), e07.

15. Paquette-Warren, J. & Harris, S. DEFINE user-guide [Internet]. The National Diabetes Management Strategy; 2015. pp. 1–43. Available at: http://tndms.ca/documentation/DEFINE/DEFINE%20user-guide-PDF.pdf (last accessed 17 September 2015).

16. The National Diabetes Management Strategy. DEFINE Program: Diabetes Evaluation Framework for Innovative National Evaluations [Internet]. 2015. Available at: http://tndms.ca/research/define/ index.html (last accessed 17 September 2015).

17. Blalock, A. B. (1999) Evaluation research and the performance management movement: from estrangement to useful integration? *Evaluation*, 5 (2), 117–149.

18. Baker, G. R. (2011) The contribution of case study research to knowledge of how to improve quality of care. *BMJ Quality and Safety*, 20 (Suppl. 1), i30–i35.

19. Crowe, S., Cresswell, K., Robertson, A., Huby, G., Avery, A. & Sheikh, A. (2011) The case study approach. *BMC Medical Research Methodology*, 11, 100.

20. Harris, S., Paquette-Warren, J., Roberts, S., *et al*. (2013) Results of a mixed-methods evaluation of partnerships for health: a quality improvement initiative for diabetes care. *Journal of the American Board of Family Medicine*, 26 (6), 711–719.

21. Ontario Ministry of Health and Long-Term Care (MOHLTC). Preventing and Managing Chronic Disease: Ontario's Framework [Internet]. 2007. Available at: http://www.health.gov.on.ca/en/pro/programs/ cdpm/pdf/framework_full.pdf (last accessed 17 September 2015).

22. Paquette-Warren, J., Roberts, S. E., Fournie, M., Tyler, M., Brown, J. & Harris, S. (2014) Improving chronic care through continuing education of interprofessional primary healthcare teams: a process evaluation. *Journal of Interprofessional Care*, 28 (3), 232–238.

23. Harris, S. B., Green, M. E., Brown, J. B., *et al*. (2015) Impact of a quality improvement program on primary healthcare in Canada: a mixed-method evaluation. *Health Policy*, 119 (4), 405–416.

24. Kotecha, J., Brown, J. B., Han, H., *et al*. (2015) Influence of a quality improvement learning collaborative program on team functioning in primary healthcare. *Families, Systems and Health: The Journal of Collaborative Family Healthcare*, 33 (3), 222–230.

25. Barr, V. J., Robinson, S., Marin-Link, B., *et al*. (2003) The expanded Chronic Care Model: an integration of concepts and strategies from population health promotion and the Chronic Care Model. *Hospital Quarterly*, 7 (1), 73–82.

26. Coleman, K., Mattke, S., Perrault, P. J. & Wagner, E. H. (2009) Untangling practice redesign from disease management: how do we best care for the chronically ill? *Annual Review of Public Health*, 30, 385–408.

27. Institute for Healthcare Improvement. The Breakthrough Series: IHI's Collaborative Model for Achieving Breakthrough Improvement [Internet]. Boston, Massachusetts: Institute for Healthcare Improvement; 2003. Available at: http://www.IHI.org (last accessed 17 September 2015).

28. O'Cathain, A., Murphy, E. & Nicholl, J. (2010) Three techniques for integrating data in mixed methods studies. *British Medical Journal*, 341, c4587.

29. Bryson, J. M., Patton, M. Q. & Bowman, R. A. (2011) Working with evaluation stakeholders: a rationale, step-wise approach and toolkit. *Evaluation and Program Planning*, 34 (1), 1–12.

30. Crabtree, B. F., Chase, S. M., Wise, C. G., *et al*. (2011) Evaluation of patient centered medical home practice transformation initiatives. *Medical Care*, 49 (1), 10–16.

31. Tricco, A. C., Ivers, N. M., Grimshaw, J. M., *et al*. (2012) Effectiveness of quality improvement strategies on the management of diabetes: a systematic review and meta-analysis. *Lancet*, 379 (9833), 2252–2261.

32. Hayes, H., Parchman, M. L. & Howard, R. (2011) A logic model framework for evaluation and planning in a primary care practice-based research network (PBRN). *Journal of the American Board of Family Medicine*, 24 (5), 576–582.