



**Centre for Health Economics
and Policy Analysis**

CHEPA WORKING PAPER SERIES

Paper 21-03

A Measurement and Evaluation Roadmap to Support Social Impact Investment in Canada

November 26, 2021

Jeremiah Hurley, PhD ^{1,2,3}

Deborah Milinkovic ^{1,2}

David Feeny, PhD^{1,2}

G. Emmanuel Guindon, PhD ^{1,2,3}

Arthur Sweetman, PhD^{1,2}

Jean-Éric Tarride, PhD^{1,2,3,4}

1 Centre for Health Economics and Policy Analysis, McMaster University

2 Department of Economics, McMaster University

3 Department of Health Research Methods, Evidence, and Impact,
McMaster University

4 Program for the Assessment of Technology in Health, McMaster
University



A Measurement and Evaluation Roadmap to Support Social Impact Investment in Canada

November 26, 2021

Jeremiah Hurley, PhD^{1,2,3}

Deborah Milinkovic^{1,2}

David Feeny, PhD^{1,2}

G. Emmanuel Guindon, PhD^{1,2,3}

Arthur Sweetman, PhD^{1,2}

Jean-Éric Tarride, PhD^{1,2,3,4}

¹Centre for Health Economics and Policy Analysis, McMaster University

²Department of Economics, McMaster University

³Department of Health Research Methods, Evidence, and Impact, McMaster University

⁴Program for the Assessment of Technology in Health, McMaster University

Acknowledgements:

This work was supported by funding from Community Foundations of Canada, the Social Sciences and Humanities Research Council of Canada (Individual Partnership Engage Grant #892-2020-0020) and the Future of Canada Project at McMaster University. We would like to thank the following individuals who contributed to this work: Melanie Thomas, Sally Hough, Allison Van, Gioia Buckley, Kelli Stevens, David Johnson, Len Usiskin, Stan Yu, Don Ewles, Ahmed Abunada, Olivia Michalczuk and Leslie Kwok. The views expressed in this report are those of the authors and should not be taken to represent the views of the Social Sciences and Humanities Research Council, Community Foundations of Canada, The Future of Canada Project, or McMaster University.

TABLE OF CONTENTS

I.	INTRODUCTION	1
II.	THE OUTCOMES-BASED PAYMENT MODEL OF IMPACT INVESTING	2
III.	MEASUREMENT AND EVALUATION CHALLENGES	3
IV.	ROADMAP DEVELOPMENT AND METHODS	3
V.	THE MEASUREMENT AND EVALUATION ROADMAP	4
1.	THE PROGRAM	7
2.	THEORY OF CHANGE – CONCEPTUALIZING THE PATHWAY FROM INPUTS TO OUTCOMES	9
2.1	LOGIC MODEL	10
	<i>2.1.1 Outputs vs. Outcomes vs. Impacts</i>	<i>11</i>
	<i>2.1.2 Reference Level and Threshold</i>	<i>11</i>
3.	MEASUREMENT AND EVALUATION	12
3.1	CONTEXTUAL ISSUES	12
	<i>3.1.1 Stakeholder Involvement</i>	<i>12</i>
	<i>3.1.2 Constraints</i>	<i>12</i>
	<i>3.1.3 Enabling or Hindering Factors</i>	<i>13</i>
	Individual-level contextual	13
	Community-level contextual factors	13
	Structural contextual factors	14
	<i>3.1.4 Reference Data, Program Size, and Threshold Contextual Issues</i>	<i>14</i>
	a. Missing Reference Data	14
	b. Program Size	14
	c. Thresholds	14
3.2	EVALUATION APPROACHES	15
	<i>3.2.1 Stage of Program Development and Implementation</i>	<i>16</i>
	a. Formulative Evaluation	16
	b. Process Evaluation	16
	c. Summative Evaluation	17
	d. Rapid (Real-time) Evaluation	17
	<i>3.2.2 Quantitative versus Qualitative Designs</i>	<i>17</i>
	a. Quantitative Methods	18
	b. Qualitative Methods	18

3.2.3	<i>Descriptive Approaches</i>	19
a.	Community-based, Participatory Designs	19
3.2.4	<i>Assessing Effectiveness: Establishing Causal Relationships Between a Program and Outcomes</i>	20
a.	Experimental Designs based on Randomization	21
b.	Observational, Quasi-experimental Designs	22
c.	Before-after Designs	22
d.	Cross-sectional Designs	23
e.	Designs that Utilize both Cross-sectional and Time-series Variation	24
f.	Contribution Analysis	25
3.2.5	<i>Mixed Methods</i>	25
3.2.6	<i>Evaluation Methods Summary</i>	27
3.3	METHODS FOR ASSESSING VALUE FOR MONEY	28
3.3.1	<i>Approaches that Do not Assign Value to Outcomes</i>	28
3.3.2	<i>Approaches that Do not Assign Value to Outcomes</i>	29
3.3.3	<i>Approaches that Assign Monetary Value to Outcomes</i>	29
	Financial Analysis of Cost Savings	29
	Cost-benefit analysis.....	30
3.3.4	<i>Social Return on Investment</i>	31
3.4	OUTCOMES AND INDICATORS	31
3.4.1	<i>Types of Indicators</i>	45
a.	Quantifiable	45
b.	Non-Quantifiable.....	45
c.	Objective	46
d.	Subjective	46
e.	Self-Reported	46
f.	Third-Party Reported	46
3.5	DATA TYPES AND ACCESSIBILITY	46
3.5.1	<i>Primary Data</i>	46
3.5.2	<i>Secondary Data</i>	47
4.	DID THE PROGRAM ACHIEVE THE SPECIFIED THRESHOLDS FOR THE OUTCOMES?	48
VI.	CONCLUSION	48

FIGURES AND TABLES

Figure 1: A Measurement and Evaluation Roadmap to Support Outcomes-based Social Impact Investment in Canada.....6

Figure 2: Interconnectedness of Outcomes31

Table 1: Common Outcomes33

Executive Summary

Social finance offers innovative ways to finance solutions to many of society's most challenging problems by attracting private investments that create both financial returns for investors and desired social and environmental impacts. Community Foundations of Canada (CFC) is working with sector partners to develop a social-finance platform called Outcomes Canada (OC) to facilitate community-driven social finance. Many community initiatives face significant capacity challenges and resourcing gaps. At present, community initiatives bear the burden of navigating a highly decentralized financing ecosystem to secure funding and other support needed to build their capacity and achieve desired outcomes. At the same time, governments, social financiers, philanthropists, and academics seek to find and support community initiatives that are producing results. This gap reflects the need for new tools, innovative financing vehicles and streamlined approaches to support community-driven outcomes in Canada.

The Outcomes Canada model aims to address these challenges by creating a multi-sector initiative through which community solutions will produce positive measurable outcomes, such as economic well-being, skills training and job creation, improved health, or other social and environmental outcomes. The model consists of four phases: (1) Sourcing and Intake—community initiatives are sourced, vetted and moved through an intake process to identify needs and opportunities; (2) Enrichment—community initiatives work with developmental partner(s) to enhance their capacity to develop their business plan; (3) Measurement and Evaluation—researchers utilize measurement frameworks to identify outcomes and estimate the outcomes achieved; (4) Buyers' Table and Social Financing—financing is secured and outcomes agreements are finalized through a process that includes outcomes buyers and social finance intermediaries. This report focuses on phase 3, Measurement and Evaluation.

Outcomes Canada will facilitate outcomes-based payment models for impact investment by connecting community organizations seeking financing for promising community-led projects; investors seeking to invest in worthy projects; and buyers (e.g., government, philanthropic organizations, corporations) seeking to achieve priority social and environmental outcomes. The success of Outcomes Canada will depend in part on the measurement and evaluation of whether a funded community project achieves the social outcomes designated in the investment contract.

The overarching goal of this collaboration between the McMaster-based research team and CFC has been the development of a roadmap to guide measurement and evaluation within the Outcomes Canada impact investment platform. Development of the roadmap required the identification of: (1) distinct components of effective measurement and evaluation; (2) measurement and evaluation methods and approaches well-suited to outcomes-based payment models; (3) indicators and data sources for common outcomes for the types of

investments targeted by Outcomes Canada; (4) next steps for developing a full, robust measurement and evaluation framework. While informed by the broader context of social finance, the roadmap addresses measurement and evaluation needs specific to outcomes-based payment models of impact investing. The roadmap incorporates as case studies two community-based programs that are referenced throughout to illustrate a range of issues encountered. Both case studies are still in their development phase and serve as good examples of community-based programs with the types of outcomes of interest to an outcomes-based payment model such as Outcomes Canada. The report uses these case studies to illustrate how issues such as conflicting priorities, resource constraints, challenges of scale, and other factors figure importantly in the design of evaluation studies an Outcomes Canada platform will demand.

Key components of the framework include:

1. The program: A community organization must develop a detailed description of the program it proposes to implement including why it believes that the program will produce the desired outcomes. This lays the groundwork for the negotiation of the Outcomes Canada contract and defines explicitly what is to be evaluated.
2. Theory of change: The design of any program embodies, either implicitly or explicitly, a theory of change, often in the form of a logic model. The logic model links program features and resources to the expected short-, medium- and long-term outcomes that the evaluation seeks to measure.
3. Measurement and evaluation: An evaluation approach to determine whether the program achieves the desired outcomes. The evaluation approach must reflect a range of contextual issues that shape what is possible for an evaluation, and includes the precise evaluation methods to be used, indicators to represent the outcomes of interest, and data by which these indicators will be measured.
4. Achievement of thresholds: The objective of an outcomes-based payment evaluation is to determine whether program outcomes reached the agreed-upon thresholds to trigger payment. In some cases, a highly successful program will have exceeded a threshold unequivocally and in others an ineffective program will have unequivocally failed to do so. In many cases, however, the answer will not be straightforward and may require further analysis.

Contextual issues that will shape the measurement and evaluation approach include the following.

- *Stakeholders*, including the community organization, the investor(s), the outcomes buyer(s), the evaluation team, program recipients, the broader community or other organizations involved in some way. Stakeholder involvement can range from simply providing funding to full participation in activities such as surveys, data collection, design and implementation of the evaluation plan including selecting outcomes,

thresholds, indicators, and the interpretation of results. The depth and breadth of stakeholder participation will impact the time and resources necessary to conduct the evaluation and the validity, reliability, and transferability of the evaluation results.

- *Constraints* include factors such as financial resources, time, knowledge and expertise, data access, and others. Importantly, many community organizations lack the capacity to undertake evaluations or even, at times, to engage well with third-party evaluators.
- *Enabling and hindering factors*: Individual-level (e.g., participant characteristics), community-level (e.g., community characteristics) and structural (e.g., organizational networks), and location-specific factors can either enable or hinder both a program's implementation and success as well as its evaluation.

Evaluation Approaches

For evaluation, the roadmap distinguished broadly between descriptive approaches, approaches that seek to establish program effectiveness, and mixed-methods approaches. Cutting across these three evaluation types, the roadmap further describes two common distinctions in evaluation: that based on stage of program implementation and quantitative versus qualitative designs. The former of these two cross-cutting classifications distinguishes:

1. *Formative evaluation* determines if a program is feasible.
2. *Process evaluation* determines whether a program is being implemented as intended.
3. *Summative evaluation* establishes the extent to which an outcome is achieved.

The latter distinguishes the degree of quantification of the outcome measures:

1. Quantitative approaches are numerical-based and measure the magnitude of a change that has occurred in an outcome.
2. Qualitative approaches are non-numerically based and give stakeholders a voice by soliciting their feelings, attitudes and beliefs about an outcome and are well suited for understanding a participant's lived experience.

With regards to the three primary evaluation approaches, descriptive approaches seek to document and describe what is happening under a program, e.g., who a program reaches, how many it reaches, and outcomes observed including the experiences of program participants. Given the emphasis of Outcomes Canada's platform to community-led approaches, descriptive, community-based designs, including participatory approaches may be important for at least some aspects of program evaluations. Crucially, because descriptive analyses do not control for other factors that could account for any differences observed between the program and other settings, they cannot be used to assess whether a program is effective at generating the sought-after outcomes.

Given the outcomes-based payment model of Outcomes Canada, full evaluations will of necessity include approaches designed to assess a program's effectiveness in producing the desired outcomes. This requires establishing a causal connection between a program and the

outcomes observed. Most suitable for most community programs envisioned for Outcomes Canada will be quasi-experimental approaches based on before-after, cross-sectional or before-after, cross-sectional designs. Finally, mixed-methods approaches combine in complementary ways different methodological approaches in a single evaluation, seeking to gain the advantages of each method and thereby strengthening the overall study design.

Mixed-methods approaches may be particularly important for the Outcomes Canada platform given its twin goals of community-led, community-engaged processes and the evaluation of program outcomes. Descriptive, community-engaged, often participatory, approaches are well suited to identifying promising programs, defining program outcomes and shaping the evaluation design, and more generally in giving voice to community organizations and community members (especially program participants). Effectiveness approaches best serve the need to establish causal connections between a program and the outcomes measured as part of determining whether a program has achieved the designated threshold levels, which is central to Outcomes Canada's outcomes-based payment model of social finance. For some settings and designs, the demands of these two methodological approaches will be in tension. Community-engaged, participatory approaches risk the evaluators becoming part of the intervention—an additional program component specific to the evaluation context that would be absent when a successful program is delivered more broadly. From the perspective of establishing causal connections between a program and outcomes, this phenomenon “contaminates” the evaluation because it precludes establishing the impact of the program alone, absent evaluator effects. These tensions will require very careful consideration of how the two designs may interact in mixed-methods evaluation in ways counterproductive to the ability to gain the advantages of each, and instead even compromising the strengths of each. This may require differential timing of the different components, applying the different approaches to different subsets of study participants, or modifying the approaches in ways that limit the negative interactions while ensuring the core evaluation objectives can still be met.

Methods for Assessing Value for Money

Some contracts will specify outcomes beyond effectiveness alone to include evidence of value for money—e.g., are the outcomes achieved sufficiently large in relation to the resources required to produce them? Determining value for money adds considerable complexity to an evaluation and makes substantially greater data demands. An important distinction among the value for money methods is whether the method assigns a value to the outcomes, and if so, whether that value is expressed in non-monetary or monetary terms.

Did the program achieve the specified thresholds for the outcomes?

In some instances, the program will unequivocally achieve the predetermined outcome thresholds, triggering payment; in others the program will not, so no payment will occur. In others still, the results will be mixed in some way (e.g., uncertainty as to whether thresholds

were achieved, or some thresholds met and others not), which will require intermediate actions as negotiated in the contract (e.g., pro-rate the overall payment to reflect partial success).

Key Conclusions

The roadmap identified critical components of any evaluation and options to consider for each, but it cannot provide specific recommendations because each evaluation must be customized to the evaluation setting. Outcomes Canada's twin goals of incorporating community-led and community-engaged approaches to program development with rigorous evaluation of program outcomes will at times come into tension. Attaining both goals will often require mixed-methods evaluation approaches using innovative designs, recognizing that such mixed-methods approaches will also be more resource intensive.

This roadmap is a first step for measurement and evaluation to support the Outcomes Canada platform. Outcomes Canada will need to develop the roadmap more comprehensively, which will require it to define more clearly the types of programs it seeks to support and develop and likely expectations of the different stakeholders integral to its model, and identify a few basic models of measurement and evaluation well-suited for such contexts. Finally, Outcomes Canada will have to make important decisions about how much and what types of measurement and evaluation infrastructure it will build in-house and what it will contract with external parties.

I. Introduction

Social finance offers innovative ways to finance solutions to many of society's most challenging problems by attracting private investments that create both financial returns for investors and social and environmental impacts. Community Foundations of Canada, a national network of 191 philanthropic community foundations, is working with sector partners to develop a social-finance platform, called Outcomes Canada, designed to mitigate the challenge of connecting potential investors with promising community-based initiatives.¹ Many community initiatives face significant capacity challenges and resourcing gaps. At present, community initiatives bear the burden of navigating a highly decentralized funding ecosystem to secure funding and other support needed to build their capacity and achieve desired outcomes. At the same time, governments, social financiers, philanthropists, and academics seek to find and support community initiatives that are producing results. This gap reflects the need for new tools, innovative financing vehicles and streamlined approaches to supporting community-driven outcomes in Canada. The Outcomes Canada model aims to address these challenges by creating a multi-sector initiative through which community solutions will produce positive measurable outcomes, such as economic well-being, skills training and job creation, improved health, or other social and environmental outcomes.

Outcomes Canada will facilitate a particular form of social finance—outcomes-based payment models for impact investment— by connecting community organizations seeking financing for promising community-led projects; investors seeking to invest in worthy projects; and buyers (e.g., government, philanthropic organizations, corporations) seeking to achieve priority social and environmental outcomes (hereafter, “social outcomes”).² Among the essential functions of the Outcomes Canada platform is the early-stage assessment of a project's potential impact (and therefore attractiveness to social investors), and then, among those successfully capitalized projects, the measurement and evaluation of whether a funded community project achieved the social outcomes designated in the investment contract.

The measurement and evaluation of outcomes is critical to Outcomes Canada's ultimate success. Measurement and evaluation for impact investment is challenging, as attested to by the numerous recent calls for the development of new measurement approaches to support meaningful and rigorous outcomes assessments.^{3,4,5,6,7,8} In the context of the Outcomes Canada platform, such challenges include those common to impact investing more generally— data consistency, quality and access;^{3,8,9,10,11} organizational capacity;^{3,8,9,11,12} and simple, relevant, rigorous and useable metrics,^{3,13,14}— and some challenges unique to the outcomes-based investment models promoted by Outcomes Canada, including the specification of quantifiable benchmarks negotiated among the parties to the outcomes-based payment contracts and multi-outcome metrics that map from outputs (what organizations do, e.g., skills training) to outcomes (e.g., employment).

The overarching goal of this collaboration between the McMaster-based research team and Community Foundations of Canada has been the development of a roadmap to guide measurement and evaluation within the Outcomes Canada impact investment platform. Development of the roadmap required the identification of: (1) distinct components of effective measurement and evaluation; (2) measurement and evaluation methods and approaches well-suited to outcomes-based payment models; (3) indicators and data sources for common outcomes for the types of investments targeted by Outcomes Canada; and (4) next steps for developing a full, robust measurement and evaluation framework. While informed by the broader context of social finance, the roadmap addresses measurement and evaluation challenges specific to outcomes-based payment models of impact investing.

II. The Outcomes-based Payment Model of Impact Investing

Social finance takes many forms, all of which share the common goal of attracting private capital to drive needed social innovation and change.¹⁵ Impact investing is one of the most prominent forms of social finance and itself includes multiple variants on a spectrum from low interest loans and recoverable grants to equities (stocks), fixed income assets (bonds) and venture capital^{16,17}

Outcomes Canada employs an outcomes-based payment model of impact investing that distinguishes itself from pay-for-performance models such as social impact bonds through its commitment to and strong support for community organizations whose initiatives, social outcomes, voice and objectives are prioritized over those of the investor. The model consists of four phases: (1) Sourcing and Intake—community initiatives are sourced, vetted and moved through an intake process to identify needs and opportunities; (2) Enrichment—community initiatives work with developmental partner(s) to enhance their capacity to develop their business plan; (3) Measurement and Evaluation—researchers utilize measurement frameworks to identify outcomes and estimate the outcomes achieved; (4) Buyers' Table and Social Financing—financing is secured and outcomes agreements are finalized through a process that includes outcomes buyers and social finance intermediaries. This report focuses on phase 3, Measurement and Evaluation.

Outcomes Canada's outcomes-based payment model will connect a community organization that seeks financing to implement a program designed to produce defined social outcomes; an investor that seeks to finance an activity that will generate both the defined social outcomes and a financial return; and a buyer who seeks to support the implementation of effective programs that achieve defined social outcomes. Under this model, the community program itself is not expected to produce a financial return. Instead, successful community programs produce social outcomes that trigger payment from the buyer to the investor, thereby generating investor return. The investment involves a contract whereby in return for investing in the community program, should the program achieve outcomes defined in the contract (i.e., meets its outcome targets, such as reducing homelessness), the buyer makes a defined

payment to the investors to provide a financial return. If the program fails to achieve its outcome targets, no payment transpires and the investors lose their investment. A project designed to achieve multiple types of social outcomes, such as reduced homelessness and increased employment, may achieve some specified targets but not others. In such cases, investors would be repaid on a prorated basis corresponding to outcomes that were achieved. Because the payment depends on achieving specific thresholds for each of multiple social outcomes, the outcomes-based payment approach creates unique demands for evidence, measurement, and evaluation compared to the predominant social-enterprise model noted above.

III. Measurement and Evaluation Challenges

The Rockefeller Foundation has noted that “impact investing” could be reduced to a simple marketing tool if “a certain level of rigour in impact measurement is not established.”^{4, p.7} A number of recent reviews articulate the measurement and evaluation challenges associated with impact investing.^{3,5,6,7} Major challenges include a tension between standardized approaches that facilitate comparison by investors of multiple investment options, and the need for outcome measures specific to the nature of each investment opportunity. Further, social enterprises are heterogeneous^{9,13}— they differ in mission, scale and scope,⁷ track and report different outcomes,³ employ different methods, metrics and assumptions,^{10,12} and often lack organizational resources and capacity to conduct rigorous evaluations.^{3,8,9,11,12} These reviews call for research on new measurement approaches that can measure outcomes in a meaningful and rigorous way to inform real-time decision-making.^{3,4,5,6,7,8}

While many similar issues arise for outcomes-based payment arrangements, their importance and nature differ and other additional challenges arise. For outcomes-based payment investing, social impact is the primary purpose of the funded program, and it is the sole criterion for determining whether an investor realizes a financial return. Consequently, the assessment of social outcomes achieved must be more rigorous and exacting. Ultimately, the evaluation must determine whether the outcomes reach the threshold to trigger the transfer payments from buyers to the investors.

IV. Roadmap Development and Methods

To develop the roadmap, we adopted a multi-faceted approach with four key elements.

- a. Consultations: Development of the measurement and evaluation roadmap required consultations with key stakeholders involved in impact investing.

b. Case-Grounding: The roadmap must address practical measurement and evaluation challenges faced “on-the-ground.” In developing the roadmap, we drew on two community programs (or “cases”) of interest to Community Foundations of Canada that are similar in nature to programs to be considered for the Outcomes Canada platform. Both programs have the potential to achieve outcomes that align with CFC’s goal of producing positive measurable social and environmental outcomes. Although the two programs are currently in their development phase, they provide specific contexts in which to assess how a roadmap can integrate principles of good measurement and evaluation while accommodating the realities of such programs.

c. Data Development and Mapping:

One objective of the roadmap is to identify outcomes that are likely to be common to multiple programs targeted by Outcomes Canada, and to identify data sources and propose methods to measure and value such outcomes within an outcomes-based payment model.

d. Conceptual Integration: Lastly, all these elements are brought together and integrated into a practical, conceptually coherent, and methodologically informed roadmap that serves the needs of CFC and the Outcomes Canada platform. This will also inform further work required to fully develop and implement a model for robust, rigorous measurement and evaluation within CFC’s Outcomes Canada platform. This required a combination of thinking, reflection, and intellectual exchange among team members and between the team and key stakeholders.

V. The Measurement and Evaluation Roadmap

Measurement and evaluation is a systematic assessment of the design, implementation and results of a program for the purposes of learning or decision-making.¹⁸ An evaluation can serve one or more purposes. It can: (a) assess whether or to what extent a program achieves the

Case Study 1

BUILD UP Saskatoon (BUS) is a Quint Development social enterprise. Quint has helped connect community members to housing, employment and other services for 25 years including resume development, setting up bank accounts, on-the-job training and now job experience through BUILDUP Saskatoon.

BUILD UP Saskatoon provides employment to community members escaping gang association or who were recently incarcerated. The participants gain experience and skills in the construction industry while renovating and retrofitting existing community housing for the social enterprise Quint. Outcomes of interest include reduced contacts with the justice system and reduced justice costs.

specified outcomes; (b) provide insight regarding how the program achieved desired outcomes to further learning and improve program design and delivery over time; and (c) give voice to various stakeholders such as funders, program administrators, program recipients, community members and others.^{10,13,19,20}

Measurement and evaluation is an iterative process—a series of questions and answers that starts with the simple but sometimes difficult question: “*what does this evaluation seek to demonstrate?*” The objective(s) of the evaluation must be clearly and explicitly articulated and accompanied by precise and actionable evaluation questions formulated with and agreed to by relevant stakeholders. For an outcomes-based payment model such as the Outcomes Canada Framework, the primary objective will be assessing performance— assessing the extent to which the program achieved desired

outcomes over a specified period. This, however, does not preclude additional secondary objectives pertaining to processes of implementation, stakeholder perceptions and experience of a program, key channels by which outcomes were achieved, and other aspects of a program and its impacts. For programs in their early development such as the two case studies mentioned above, such learning would certainly be a valuable objective.

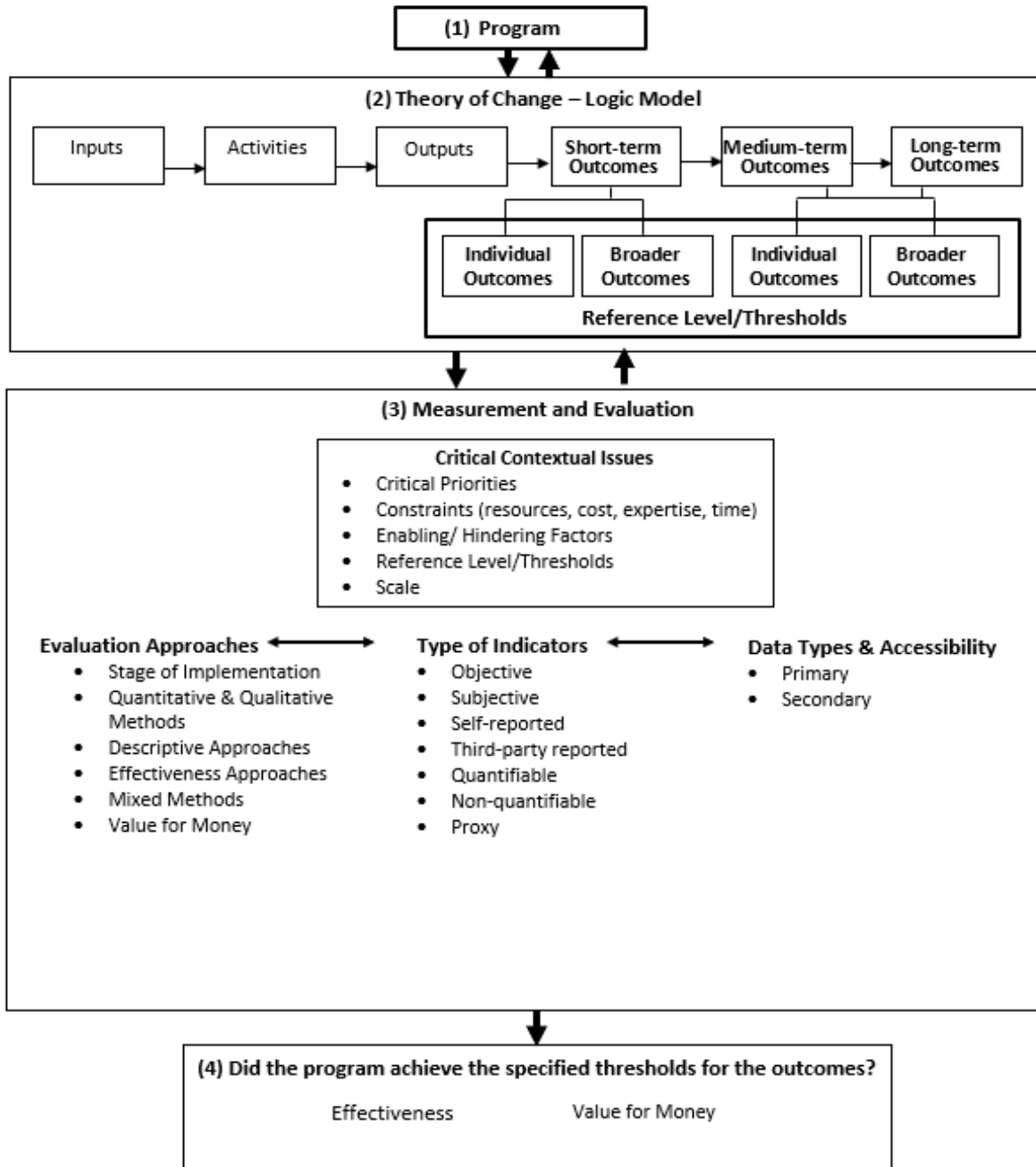
Figure 1 (below) is a schematic representation of the measurement and evaluation roadmap. At the top is the program^a being evaluated. The design of any program embodies, either implicitly or explicitly, a theory of change depicted in the figure as a logic model. The logic model links program features and resources to the expected short-, medium- and long-term outcomes²¹ that the evaluation seeks to measure.

Case Study 2

SALSA is a Spence Neighbourhood Association social enterprise in Winnipeg that provides employment opportunity to underserved youth from the Spence neighbourhood. Participants gain experience and skills in agriculture, food preparation, marketing and food distribution. The priority outcomes for SALSA are youth employment and revenue gained through SALSA sales. Program revenues will be used to support community gardens in a neighbourhood with few fresh food options and support youth wages. The outcomes from this program include better health through better nutrition and improved quality of life.

^a We use the term “program” to represent the set of activities the community organization proposes to implement. In the literature, this set of activities is referred to, among other terms, as a project, a program, an initiative, an intervention, a treatment. For clarity we use the term program consistently in the roadmap.

Figure 1: A Measurement and Evaluation Roadmap to Support Outcomes-based Social Impact Investment in Canada



From this flows the design of the measurement and evaluation work, which itself includes many components. This embodies a range of design decisions that reflect a set of contextual factors—priorities, constraints, enabling factors—that shape what is possible for an evaluation, decisions regarding the precise evaluation approach(es) to be used, indicators to represent the

outcomes of interest, and data by which these indicators will be measured. While evaluation encompasses all these elements, *measurement* refers more specifically to the process of ascertaining the presence, amount, or extent of a variable of interest for the study. Measurement can be achieved to varying degrees of quantification, to varying degrees of accuracy, and to varying degrees of precision. It can be purely qualitative (presence/absence) or quantitative, striving to estimate the degree to which an outcome has been achieved. The most suitable approach depends on the evaluation objectives and the outcomes.

Finally, measurement and evaluation enables one to answer the fundamental question for outcomes-based financing — whether the program achieved the threshold levels of outcomes as specified in the investment contract, and if appropriate, place a value (often, but not always in monetary terms) on those outcomes achieved.

Importantly, while the figure depicts these components or stages as occurring sequentially, interdependencies among the components mean that, in reality, the process is highly iterative, often circling back and forth among them throughout the overall process. We now turn to a more detailed discussion of each component. The sub-section numbers in the roadmap text below correspond with the numbering of the components in the roadmap Figure 1.

1. The Program

The community organization must develop a detailed description of the community program. This description must state clearly the activities associated with the program, the resources the program will use, the specific outcomes it expects to produce, and why it believes that the program will produce these outcomes. This program description provides the foundation for establishing the outcomes that will trigger payment and the evaluation approach to be employed.

Participating organizations may put forth either a pre-existing program seeking alternative sources of funding or a new program seeking to address gaps in environmental or social services needs in their community. These two situations create distinct evaluation challenges. For example, it may be more difficult to establish a reference “no program” level of outcomes against which to compare outcomes under the program.

Many community organizations will face challenges mounting a new program or even an expanded version of an existing program. A unique feature of the Outcomes Canada platform is its commitment to capacity building for those community organizations participating in the platform.

Although not explicitly depicted in the figure, establishing the outcomes-based payment contract for a program requires negotiation among the community organization, investors, and

buyers, a process that, given Outcomes Canada’s emphasis on community-oriented, community-led initiatives, will often also include community members, and other relevant community-based stakeholders. Negotiation is necessary because, although the parties have shared interests, they may weigh them differently, have other interests and priorities not shared by all parties, and have different commitments in this overall process. While the need to negotiate the outcomes and their associated target thresholds may be self-evident, the negotiation will also encompass elements of the measurement and evaluation approach to be used. Choices over these elements will reflect the nature of the evidence required, will privilege some voices over others, will serve the interests of the partners differently, and so forth. As a social finance platform designed to support community-based initiatives and outcomes, it will be important that the negotiation process adopted by Outcomes Canada prioritize the voice and perspectives of the community organization in the process.

Negotiation also does not end once the contract is signed. Program implementation and evaluation seldom unfold precisely as anticipated, requiring that choices and issues be revisited as the evaluation proceeds. Indeed, a framework for handling such contingencies should be negotiated as part of the contract and that framework should include the evaluation team as part of the process—once the program is launched and the evaluation is underway, the evaluation team itself becomes a stakeholder/partner.

Four points bear emphasizing. First, rather than arms-length, these arrangements are best viewed as a type of partnership that requires regular engagement by the partners. Second, as noted, the iterative nature of the overall process will require that early decisions be revisited,

Prioritization

BUILD UP Saskatoon (BUS) has identified reduced contact with the justice system and reduced costs for the justice system as their primary outcomes of interest and wants to engage in an outcome-based model to raise the funding to achieve this goal. The model engages with three parties.

Community Organization: BUILD UP Saskatoon

Investor: Saskatoon Community Foundation

Buyer: Saskatoon Police Service

BUILD UP Saskatoon prioritizes reduced contact with the justice system because they view it as a pathway for achieving other social benefits such as increased quality of life for their participants and family stability. The Saskatoon Police Service prioritizes cost savings because savings can be re-deployed to achieve other objectives of the police services. Although both objectives are potentially achievable, cost saving will take longer to manifest, requiring an evaluation approach that is more resource intensive.

and other modifications be made as implementation and evaluation proceed. Third, all good evaluation rests on a foundation of stakeholders having a clear understanding of and consensus on the objective(s) of the evaluation. These objectives must be translated into unambiguous, empirically answerable evaluation questions. Failure to do so can lead to poor choice of evaluation design, to challenges interpreting study results, and to avoidable conflict among stakeholders. Fourth, while some issues encountered are, in a sense, purely technical (e.g., will this randomization method lead to true randomization?), many seemingly technical matters carry with them implicit value judgements or assumptions (e.g., processes for how disparate outcomes are aggregated; the process by which outcomes are valued) and many technical choices have implications for non-technical matters of value for stakeholders (e.g., how the approach to community engagement affects community members' ability to give voice to their perspectives). This intermixing of the technical and the normative is inescapable, must be recognized, and needs to be addressed when making design choices and negotiating the contract and the evaluation approach.

2. Theory of Change – Conceptualizing the Pathway from Inputs to Outcomes

A theory of change establishes the conceptual underpinnings of the causal pathways that link program activities and resources (inputs) to the sought-after outcomes: what is the logic or reasoning that connects them? The theory of change is often informed by empirical evidence regarding the extent to which these hypothesized causal pathways have been empirically verified and the confidence in that evidence. To the extent that such evidence exists, it is useful to cite it in support of the theory of change.

Although the hypothesized causal linkages are often implicit for the community organization, best practice calls for an evaluation to articulate explicitly the underlying theory of change for a program. Doing so helps reveal assumptions underlying the design of a program and clarifies and sharpens the reasoning linking the program to the outcomes. It informs the design of the evaluation and the required measurements to ensure that, to the extent desired and feasible, the evaluation demonstrates not only whether the program works but why it works. This information can be essential for scaling up the size of a program in the original sites and for introducing the program in new locations. The theory can identify intermediate outcomes that offer the potential to assess effectiveness in those situations when ultimate outcomes are either hard to measure or take considerable time to be realized. It enables potential investors to understand the rationale for the program design and give them greater confidence that a program genuinely has the potential to achieve the sought-after outcomes. This transparency can also aid negotiations of the impact-payment contract between the community organization, the investor, and the buyer.

2.1 Logic Model

The theory of change is often represented visually by a logic model, such as that depicted in Figure 1.^{13,19,22,23,24} The key components of a logic model include inputs, activities, outputs, and outcomes.

Inputs are the resources required to implement and administer a program. Resources include financial, human (staff, volunteers, external consultants, etc.), materials and equipment, office space and others dependent upon the goals and activities of the program.

Inputs are used to produce *activities*— the things that organizations do as part of the program. These activities could include workshops, training sessions, skills development, counselling, and others depending upon the program and its objectives. For example, BUILD UP Saskatoon provides assistance with resume development, setting up bank accounts, and on-the-job training and job experience such as framing, drywalling, plumbing, electrical, finishing, painting, fencing, and landscaping.

Activities produce *outputs*— the goods or services that a program produces. The output can be thought of as the necessary tools or intermediate phenomena required to achieve a desired outcome and is typically expressed as a count.²⁵ For example, the activities for BUILD UP Saskatoon include the number of employees trained, consistent weeks worked per employee, bank accounts set up, and new certifications obtained.

Outputs then produce *outcomes*—the ultimate phenomena of interest. Outcomes can be classified in a number of ways.

Common distinctions include, for instance, the length of time over which they arise, e.g., short-term, medium-term, long-term; the unit to which they accrue, e.g., an individual, the family or friends, the community, the government, society; their type, e.g., educational, employment-related, health-related, environmental; whether they are monetary (e.g., expenditures saved) or non-monetary (e.g., improved well-being). How they are best characterized varies from

Simple Logic Model Example

Theory of Change: Employment research suggests that employment is associated with reduced contact with the justice system, but this effect is conditional on access to stable quality employment versus any employment.⁷⁵ BUILD UP Saskatoon's aim is to provide the target population (previous gang members and ex-prisoners) with this type of quality stable employment with the objective to reduce recidivism.

Inputs: supervisor-trainer, tools, etc.

Activities: the provision of on-the-job training and work experience through BUILD UP Saskatoon.

Outputs: Number of participants trained through BUILD UP Saskatoon.

Outcomes: reduced contact with justice system (short-term, medium-term), reduced justice costs (long-term).

evaluation to evaluation and they are not mutually exclusive. Indeed, it can be useful to consider the intersection of their characteristics, e.g., short-term education gains versus long-term savings of government expenditures. Short- and medium-term outcomes often lead to the longer-term outcomes, but not always, which is why it is important to understand the mechanisms that underpin the theory of change.

Outputs vs. Outcomes vs. Impacts

The literature on impact measurement and evaluation for social finance includes some confusing terminology regarding outputs, outcomes, and impacts. Different frameworks use the terms differently, and one concept may be labelled an output in one but an outcome in another; similarly, some frameworks define impacts as long-term outcomes whereas others define impacts as societal-level changes. Because there is neither a single correct way to define or classify the underlying phenomena nor a dominant convention, the roadmap adopts the following usage.

First, the roadmap distinguishes outputs and outcomes. As mentioned previously, outputs are the intermediate goods, services or attributes that contribute to achieving the ultimate desired outcomes of interest. Developing a resume, setting up a bank account and obtaining identification are all necessary program outputs for achieving an employment outcome for a program participant, but they are not outcomes. The distinction between outputs and outcomes can be important because outputs do not necessarily translate to outcomes – e.g., the number of program graduates, resumes created, or job interviews held does not equate to the number of people ultimately employed, although they are often necessary steps toward achieving that ultimate outcome.

Second, the roadmap does not distinguish outcomes and impacts but rather refers simply to outcomes and distinguishes as appropriate short-term, medium-term, and long-term^b outcomes; or individual, community, societal outcomes.

2.1.1 Reference Level and Threshold

“Reference level” in Figure 1 refers to the level of an outcome in the absence of the program. It is an essential element of an evaluation to assess the extent to which a program improves outcomes and, in particular, whether the program improves outcomes more than specified by the defined outcome threshold(s). For a new program, the reference level is often the baseline level prior to introduction of a program. For an existing program, the reference level may be that observed in a similar setting without the program. In either case, establishing this reference outcome level requires collecting relevant data from a context without the program.

^b Exactly how short-, medium-, and long-term are defined will depend importantly on the program under consideration; there is no fixed duration associated with each. But as a rough guide, one can think of short-term as less than two years, medium as two-four years, and long-term as more than four years.

The threshold refers to the target level of the outcome a program must achieve to trigger the payment to an investor under the investment contract. A threshold can be specified as an absolute level of the outcome to achieve or as a specific amount of difference compared to the reference level. The level at which a threshold is set is usually based on the best available evidence from similar programs regarding what a program can be expected to achieve.

3. Measurement and Evaluation

3.1 Contextual Issues

Contextual issues will drive key features of the evaluation design including stakeholder involvement, evaluation approaches and methods, measurements and indicators, data collection, and related issues. Common contextual issues include contract- and program-related factors such as resource and time constraints, the nature and scale of a program, and broader environment-based factors such as population characteristics, community characteristics, data availability, and stakeholder priorities, knowledge and commitment.

3.1.1 Stakeholder Involvement

Negotiations of payment-based outcome-contracts must define the roles of the key stakeholders. Stakeholders include the community organization, the investor(s), and the outcomes buyer(s), the evaluation team, community members including those directly affected by a program, and other relevant organizations. Stakeholder involvement can range from simply providing funding, to identification of program outcomes and indicators and setting thresholds, to participating in broader aspects of the evaluation design, to full participation in surveys and data collection, to aiding the interpretation of results. The level of stakeholder involvement will influence the type of evaluation method utilized, time required to conduct an evaluation, the cost, and other matters. These types of issues must be negotiated at the time of establishing the outcomes-based payment contract as discussed above in section 1.1.

3.1.2 Constraints

Constraints include financial resources, time, knowledge and expertise, data access and others. Importantly, many community organizations lack the capacity to undertake evaluations or even, at times, to engage well with third-party evaluators.^{3,8,9,11,12}

Resource constraints will influence choices regarding evaluation approach and methods, types of indicators, data collection and data access. For example, collecting data through surveys or focus groups often requires more resources than relying on data from third-party survey or

administrative data. Collecting primary data through customized surveys or focus groups allows greater flexibility for measuring outcomes but can be very expensive; secondary data can often be less expensive but may offer only a limited set of possible indicators (more on this below).

Time constraints influence all aspects of the evaluation. Of particular importance is the period of time over which outcomes will be measured. Some outcomes manifest in a relatively short period of time (e.g., eat more vegetables) while others can take several years to manifest (e.g., improved health). Other things equal, medium- and long-term outcomes are more costly to track and measure. This can present important challenges as short-term outcomes are often poor proxies for long-term outcomes.

3.1.3 *Enabling or Hindering Factors*

Contextual factors can have either an enabling or hindering effect on both a program's implementation and success as well as its evaluation.

Individual-level contextual factors include features such as race, gender, education, health, mental health, language, family history or past-experience. The target population will have certain characteristics that make the objectives more (less) difficult to realize.

Community-level contextual factors include features such as geographic location, economic condition, community

Constraints & Contextual Factors

Due to the early stage of the BUILD UP program, it faces several constraints.

- Small sample size
- No access to costing data
- Lack of access to police data
- Minimal internal expertise in program evaluation

Individual-level Factors: Because the program participants of BUILD UP Saskatoon predominantly identify as Indigenous (80%), best practice calls for inclusion of the Indigenous communities in the evaluation process, which will have implications for the evaluation approach, outcomes and indicators, and time and resources.

Community-level Factors: An unexpected consequence of Covid-19 is that the construction industry is experiencing a boom, which will create demand for skilled construction workers, as well as retro-fit contracts for BUILD UP. This may facilitate expansion of the BUILD UP program and enhance employment outcomes for participants.

Structural-Level Factors: Although BUILD UP has only been in operation since 2018, it benefits from established practices and community connections through Quint Development. Quint Development is the community organization responsible for launching BUILD UP Saskatoon and Quint has helped connect community members to housing, employment and other services for 25 years.

resources (housing, parks, recreational centres, grocery stores, level of equality, etc.) It may, for example, be more difficult to attain employment objectives during a recession than during a boom period, or in an economically depressed community compared to a booming community.

Structural contextual factors relate to the organizational ecosystem (interconnectedness and cooperation between organizations and networks, other organizations operating in similar space, time since the program began). It might be easier to reach certain objectives in a location with substantial community outreach if this outreach has facilitated a high level of trust and mutual respect between providers and recipients. Additionally, over time, a program may experience diminishing returns either because room for improvement has diminished or because the target population has become progressively more difficult to reach if those most adaptable to change have already been treated.

3.1.4 *Reference Data, Program Size, and Threshold Contextual Issues*

a. *Missing Reference Data*

Reference-level data are essential for most evaluation designs (e.g., any design that involves a before-after component). However, such reference data may not be available. This is especially a problem for a before-after design when the evaluation is conducted retrospectively and no data were collected in the pre-program period, or the program existed before transitioning to being supported through the Outcomes Canada platform. When reference data are absent, it may be possible to approximate a reference level through the use of secondary data—data that were collected for other purposes such as needs assessment records, school data, national survey data or stakeholder recall, etc. However, this approach could have a bearing on the validity of the evaluation. The absence or quality of reference-level data, regardless of cause, influences the choice of evaluation approaches and the methods used for analysis.

b. *Program Size*

Many evaluation designs, and quantitative designs in particular, require large numbers of program participants.^c There is no fixed minimum—such a minimum will depend on the design chosen— but often requires hundreds of participants. Achieving such scale may be difficult for some community-based programs, especially those located outside large cities. Evaluation designs will have to reflect a realistic assessment of achievable participation rates and timelines required to attain needed participant numbers.

c. *Thresholds*

Establishing realistic, relevant thresholds is essential to the success of an outcomes-based contract. If set too low or easy for an organization to achieve, investors may be paid for

^c For ease of exposition, throughout the roadmap we refer to the those participating in a program as “participants.” In most cases the most natural interpretation is that these are people but depending on the program the participants could be organizations of various types, jurisdictions, or other types of entities.

outcomes that would have occurred without any investment. If set too high, community organizations (and investors) could get discouraged and refuse to participate in outcomes-based financing. Also of importance for thresholds is the tendency for community organizations to be overly optimistic and consistently overestimate the impact a program will have on outcomes and underestimate costs.^{26,27} Thresholds should be based, therefore, on the best evidence available. Evaluators can play a critical role in both gathering such evidence and interpreting it for the context at hand.

3.2 Evaluation Approaches

Whole books have been written on basic evaluation methods. The intent of this roadmap is to provide a brief, high-level overview of the most relevant measurement and evaluation issues and concepts in the context of the outcomes-based payment model. The roadmap highlights critical choices that will have to be made when deciding on an evaluation approach. Some of these choices will be made implicitly or unconsciously; the roadmap tries to raise awareness of these choices so that they can be made more deliberately. Any given evaluation approach has many facets and, within certain limits, the options with respect to each of these facets can be combined in numerous ways to develop specific evaluation designs. Any given evaluation, for example, can focus to varying degrees on process versus outcomes, use varying amounts of qualitative and quantitative data, strive for varying degrees of rigour in establishing a causal relationship between a program and the outcomes observed, engage to differing degrees with stakeholders, and so forth. Choices will be made for each of these facets of an evaluation; the challenge is to do this consciously to identify the best combination of the choices given the objective of the evaluation. The roadmap highlights critical options and some considerations that bear on the choice among options.

Evaluation methods are constantly evolving. In recent years, for instance, various types of “rapid” evaluation (discussed below) have been developed; community-engaged, participatory approaches have become more common; methods for identifying causal relationships have expanded; and mixed-methods approaches that strive to incorporate the strengths of different methods have become more sophisticated. Many variations are possible within a given type of evaluation approach, enabling greater customization to respond to the specific needs of an evaluation.

Amidst this diversity, below we divide approaches into three broad categories: descriptive approaches, approaches that seek to determine whether a program is effective, and mixed-methods approaches. This crude categorization does injustice to aspects of various methods, but it aids in highlighting some crucial features of evaluation. One must be careful, however, not to take these labels too literally or restrictively; in reality, the methods lie within a continuum rather than in cleanly distinct categories.

Before delving into these issues, however, we briefly present two long-standing, common distinctions when classifying evaluation designs relevant to all evaluation contexts and that cut across descriptive, effectiveness, and mixed-methods approaches: (a) the stage of program development and implementation, and (b) quantitative versus qualitative designs.

3.2.1 Stage of Program Development and Implementation

Program evaluation commonly distinguishes three broad types of evaluation pertaining to different stages of program development: formulative, process, and summative evaluation. More recently, new types of rapid evaluation have emerged that blend these categories.

a. Formulative Evaluation

A formulative evaluation determines whether a program is feasible, acceptable and appropriate.²⁸ Such evaluations are used when a new program/activity is developed, or an existing program/activity is substantially modified. In the context for Outcomes Canada, a formative evaluation may be called for if there are questions as to the feasibility of the program. Because the roadmap focuses on outcomes evaluation, we assume that a formulative evaluation has already been completed.

b. Process Evaluation

Process evaluation determines whether a program has been implemented as intended. The focus of this type of evaluation is on program inputs, activities and outputs.²⁹ Process evaluation can function as a learning tool used to provide feedback to stakeholders, to gain understanding on how well the program is running, and to identify problems.²⁸ Although the focus of the roadmap is on summative evaluation (see below), many summative evaluations will include elements of process evaluation, which can inform why a program did or did not achieve the desired outcomes.

We note one particular approach, sometimes called “monitoring evaluation,” that is a process-oriented activity undertaken to improve the quality and validity of summative, outcome evaluations.³⁰ Monitoring is an ongoing process of tracking the key indicators of progress such as inputs, activities, and outputs, and occasionally outcomes, providing assurance that progress is being made in regard to the program’s objectives.^{30,31} Monitoring can reduce the end-stage burden of data collection in the outcome evaluation phase by ensuring program databases are complete and accurate, reference-level data have been collected, costing information is recorded and updated, and the integrity of participants and comparison groups is verified and confirmed (e.g., only eligible individuals have actually participated).³⁰ Monitoring can also provide investors and buyers with a greater level of assurance, through feedback, that the program is on target.

c. *Summative Evaluation*

Summative evaluation determines the extent to which a program achieves the intended outcomes.²⁹ Since this is the central evaluation challenge for outcomes-based payment finance, the roadmap focuses on summative evaluation.

In practice, use of these different evaluation approaches can be complementary and overall evaluation can incorporate elements of each, particularly process and summative evaluation. In the context of an outcomes-based payment model, process evaluation can support continued learning and ensure the program is implemented as intended and, as such, may complement a summative evaluation. Such integration should be done deliberately as part of the overall evaluation plan as the different approaches make different demands on an evaluation. Further, while such integration can produce more useful information when well done, if done poorly it can introduce sources of bias and other problems that can compromise the validity of study findings. Process evaluation, for instance, often engages a higher level of stakeholder involvement, which can make the evaluation susceptible to sources of bias.

d. *Rapid (Real-time) Evaluation*

Real-time evaluation is a relatively new type of rapid evaluation approach that deliberately integrates elements of process and summative evaluation in a particular way.^{32,33} Real-time evaluation incorporates rapid evaluations designed to quickly and systematically conduct evaluation³⁴ and, as the name implies, occurs alongside implementation of the program in real time.¹⁹ Real-time evaluation provides immediate feedback to stakeholders regarding program processes and outcomes, enabling a program to be modified to improve program delivery.^{19,32,35} The timing of the rapid evaluations corresponds to process evaluation, but some of the learnings sought correspond to summative evaluation, particularly if the results of the evaluation are time sensitive or crisis-driven such as in a pandemic or other humanitarian crisis.³⁴ Real-time evaluation fits within a broader framework of continuous-quality improvement for service delivery. It is participatory in nature^{32,34,35} and, as such, is not recommended as a primary basis for summative evaluation, for which it would be seen to suffer from challenges to reliability, validity and generalizability and to be subject to stakeholder-induced biases,¹⁹ though some of these issues can be mitigated through triangulation – employing more than one data collection method and perspective.³⁶

3.2.2 *Quantitative versus Qualitative Designs*

It is common to distinguish evaluation designs based on the degree of quantification of the outcome measures used to determine whether the program achieves the sought-after outcomes. The most common distinction is between quantitative approaches and qualitative approaches. (Many also distinguish studies that include both quantitative and qualitative components as “mixed-methods” studies, but we reserve this term to more general use and discuss it below). Quantitative methods broadly refer to methods that manipulate numerical

data to assess outcomes; qualitative refers to methods that use non-numerical data (e.g., stakeholders' feelings, opinions, and lived experience). Quantitative methods can range from calculating simple descriptive statistics such as means and medians to sophisticated statistical and econometric methods commonly based on various types of regression analyses. Quantitative analyses are not inherently more rigorous than qualitative methods as is commonly assumed. The choice of approaches must be dictated by the evaluation objectives and the nature of the outcomes.

a. *Quantitative Methods*

Not surprisingly, quantitative approaches are best suited to contexts in which the outcomes can be logically and naturally measured quantitatively. Quantitative approaches are useful when a study seeks to measure the magnitude of a change that has occurred in an outcome. Many of the methods discussed below that strive to establish causal relationships are based on quantitative data. As noted, quantitative analyses can range from the presentation of simple descriptive statistics to sophisticated statistical and econometric methods. Further discussion of specific methods goes well beyond the scope of this roadmap.

Quantitative Methods

The use of quantitative methods requires outcomes that can be measured quantitatively.

The primary outcome for the SALSA program is youth employment. Employment lends itself well to quantitative measurement along a scale from simply employed/not employed over a given period, to full-time versus part-time employment, to the number of hours worked or the earnings obtained. Each of these indicators of employment can support an overall quantitative approach to the evaluation.

b. *Qualitative Methods*

Qualitative analyses offer rich and valuable insights not possible through quantitative methods.^{37,38} These methods give stakeholders a voice by soliciting their feelings, attitudes and beliefs about an outcome and are well suited for understanding a participant's lived experience.²⁰ Qualitative methods also help evaluators identify and understand intangible factors that may influence a program's success such as the role of social norms, gender, ethnicity, religion and other contextual factors.³⁷ And, importantly, qualitative methods can help interpret quantitative results^{39,40,41,42} by giving evaluators a deeper understanding of the

^d There are differing degrees of quantification, which determines the kinds of data manipulation that make sense. Sometime numerical values (ordinal data) indicate only whether a value is greater or less than another (e.g., rate your health status as 5 = excellent, 4 = very good, 3 = good, 2 = fair, 1 = poor); others (age, income) are fully cardinal.

complex relationships between the internal and external factors that impact the program.^{42,43} Common qualitative methods include observation, focus groups, interviews, and open-ended survey questions.

3.2.3 Descriptive Approaches

Descriptive approaches seek to document and describe what is happening under a program. Descriptive approaches can document, for instance, who a program is reaching, how many it is reaching, and critical outcomes observed under the program, including the experiences of program participants. It may even provide descriptive, comparative outcomes under the program and in other similar settings.

As such, they can provide valuable, rich insight about a program to evaluators and other stakeholders. Crucially, however, because descriptive analyses do not control for other factors that could account for any differences observed between the program and other settings, they cannot be used to assess whether a program is effective (i.e., did the program itself cause the observed differences in outcomes?). Therefore, descriptive approaches will be seen as second-best when a fuller assessment of effectiveness is desired but is not feasible for reasons such as lack of reference-level data to enable program/no-program comparisons or when a program is not of sufficient scale to allow for fuller evaluation of effectiveness. But, importantly, in many cases descriptive approaches will be the method of choice for all or part of an evaluation precisely because they can provide information not otherwise possible. This is particularly true for community-based participatory evaluation approaches, which can uniquely describe, capture and document the experiences of participants and stakeholders, including some changes in their lives associated with participation. Such evidence may even be identified as an explicit outcome of interest.

a. Community-based, Participatory Designs

Community-based, participatory designs prioritize stakeholders but can vary in their depth and breadth of engagement. Stakeholders can include program participants, funders, program administrators, community members, and others. Stakeholder engagement can range from full engagement in all decision-making to that of an advisory role.^{44,45} Participatory approaches strive to give voice to those involved in a program, grounded in an ethical concern that program design and evaluation should not silence the people a program is intended to support.^{44,45,46,47}

Qualitative Methods

The use of qualitative methods requires outcomes that can be measured qualitatively.

Building on the SALSA program, youth employment outcomes can also be represented qualitatively such as job satisfaction, the impact of employment status on self-esteem, assessment of job readiness (e.g., to what extent did the program prepare you for the current employment opportunity?), the impact of employment on future goals, etc.

Such community-based participatory evaluation approaches take on an advocacy stance in evaluation design with stakeholders—and most importantly study participants—actively engaging them in the design and implementation of the evaluation.^{42,46,48,49,50} As such and by design, such participatory approaches eschew the goal of formally establishing a causal relationship between a program and outcomes, and instead prioritize multiple viewpoints, stakeholder empowerment, local knowledge, participants' lived experiences, and shared learning.^{44,46,48} Achieving these latter evaluation goals requires deep engagement by both the evaluators, program participants, and other relevant community members. As such, participatory approaches generally use qualitative methods, can be resource intensive, and require considerable time as trust is established among stakeholders. The rich information generated enables evaluators to draw implications about program design and the impacts it is having in the program setting. Some of these impacts, however, will be specific to the community under analysis and one must be careful in generalizing to other communities and settings that would experience a program differently, weigh and value experiences differently, and therefore may seek different designs.

3.2.4 Assessing Effectiveness: Establishing Causal Relationships Between a Program and Outcomes

A central goal of all outcome evaluations is to assess a program's effectiveness: does it work? That is, does the program produce the expected outcomes, where those outcomes are defined in terms of the natural units of the outcomes, e.g., jobs obtained, number of homeless, rate of depression among participants. This is a core evaluation objective for outcomes-based payment models: did a program achieve the outcomes specified in the investment contract? In most cases, this calls for establishing a causal connection between the program and the observed outcomes such that there is confidence that the outcomes observed result from the program and not some other phenomenon (e.g., the outcome was not due to another organization operating in the same space, or the outcome would have happened anyway due to another

Participatory Evaluation

The ethical conduct for research with Indigenous peoples calls for engagement with relevant Indigenous communities. Given that 80% of program participants of BUILD UP Saskatoon are Indigenous, this calls for a community-engaged approach, possibly including participants, neighbourhood leaders and the Indigenous communities from which the participants originate.

The purpose of this approach is to give voice to the participants and their communities, and to enable them to contribute to the interpretation of findings and the meaning of those findings, and more generally to understand impacts beyond the narrow focus on individual-level recidivism alone.

concurrent change). Evaluations designed to establish causation draw primarily on quantitative methods.

There can of course be other objectives and establishing such causation will not always be a primary objective, but questions of causation will arise in most evaluations. The answer to the question of whether an evaluation has established a causal relationship is not a simple yes-no. Different evaluation designs can lead to differing degrees of confidence in the extent to which an evaluation is able to establish causation.

a. *Experimental Designs based on Randomization*

Experimental designs based on randomization of study participants to the program or a control

Establishing Causality

In the context of SALSA, establishing causality means demonstrating that an increase in vegetable consumption by low-income community residents is due to increased access to vegetables grown in the community gardens funded by SALSA. The ability to demonstrate this could be threatened by other changes that coincide with the program.

For example, what if unknown to the evaluators a provincial program began running concurrently that provides low-income families with a weekly credit to purchase fresh vegetables?

Or imagine that a public health campaign has made families at all income levels more aware of the benefits of eating fruits and vegetables. As a result, people at all income levels are eating more vegetables.

Is the increased consumption of vegetables a result of the community gardens or one of these other factors?

group are commonly seen as a reference standard for establishing causation.

Randomization, if done well, ensures no unobserved differences between participants in the two groups, ensuring that measured differences in the outcomes between the two groups are caused by the program itself. Such randomized designs have high levels of internal validity—the ability of an evaluation to establish a causal relationship within the study itself.^{19,42,51}

Randomized experimental designs have important limitations.^{52,53} For example, they are difficult and highly demanding to implement, rendering them infeasible in some contexts and more generally expensive and resource intensive; in some contexts they can raise ethical issues;^{54,55,56} and the artificial conditions required to execute a randomized design well can reduce their external validity—the ability to generalize the results from the study setting to other settings.^{19,42,51} However, even when a randomized experimental design is either infeasible or overall not the best choice, the ideals grounded in such designs can serve as a reference point to assess alternative methods, especially in relation

to issues associated with establishing causation.

b. *Observational, Quasi-experimental Designs*

Because randomized experiments are often neither feasible or desirable, evaluators have developed multiple evaluation designs that compare outcomes across two groups—one made up of participants in a program and the other individuals not in a program—for which assignment to each group is not random and is due to multiple factors (often unknown to an evaluator) that determined who is in which group. Such approaches are based on observational, quasi-experimental designs, so designated because an aspect of the program design and context enables researchers to identify causal effects even in the absence of randomization. Such studies are often carried out retrospectively, exploiting a serendipitous aspect of a program. But importantly, such features can be deliberately integrated into a program or an evaluation design to enable stronger, prospective quasi-experimental studies. For example, the rollout of a large program can be staggered such that those waiting to enter the program serve as the comparison group for the evaluation.⁵⁷

Many types of quasi-experimental designs exist with the different designs offering differing degrees of confidence in the extent to which the program of interest caused the observed differences in outcomes.

c. *Before-after Designs*

A before-after design compares the program participants' outcomes after participating in a program to their outcomes before participating in a program. The participants act as their own comparison group. An advantage of before-after design is that it does not require a separate comparison group from the program participants. A disadvantage of such a design is it requires good baseline data on the participants' outcomes prior to joining a program; a before-after design is not feasible where baseline data are not available. A before-after design is relatively weak for establishing a causal relationship between program participation and post-program outcomes because other factors besides participation may have changed during the before and after periods. For example, if the broader economic conditions improve

Before-After Design

The objective of BUILD UP Saskatoon is to provide skills and employment to individuals who have had or are likely to have contact with the justice system.

Now imagine that because the evaluators have access to baseline data, they decide to compare participants' recidivism rate before and after program implementation. The evaluation indicates that the program was a success: participants had fewer encounters with the justice system after the program.

But what if, unknown to the evaluators, shortly after the program began the local police services budget was cut due to the current economic climate and these cutbacks resulted in far less active community policing?

Is an observed decrease in recidivism rates due to the program or due to the changes to police services?

(deteriorate) between the before and after periods of an employment training program, it can be difficult to distinguish how much of an observed increase (decrease) in employment is due to participation in the program and how much is due to the broader economic conditions that made it easier (harder) in general to get a job.

d. *Cross-sectional Designs*

A cross-sectional design compares the outcomes observed for individuals enrolled in a program to those of a comparison group of individuals who did not participate in the program. The assessment of the program's impact is based on the difference in outcomes observed between the two groups. An advantage of purely cross-sectional designs is that it does not require baseline data for those who participate in the program; a disadvantage is that it requires choosing an appropriate comparison group for which relevant data are available that can serve as the reference the "no program" level for assessing outcomes. The comparison group should be chosen to be as similar as possible to those who participate in the program. Even with carefully chosen comparison groups, a purely cross-sectional design is relatively weak for establishing a causal relationship between program participation and participant outcomes because there are often unobserved differences between the two groups that might account for the differences in outcomes for reasons other than the program.^e In a problem analogous to that described above for the employment

Cross-Sectional Design

Building on the previous box, now imagine that because there are no baseline data, the evaluators decide to compare outcomes for program participants to those in a community with similar demographic and socio-economic characteristics. The evaluation indicates that the program was a success: participants from the program had fewer encounters with the justice system than did the comparison group.

But, what if unknown to the evaluators the comparison group lived in a community with a "tough-on-crime" policy?

Is the difference in recidivism the result of the BUILD UP program or is it the result of the tough-on-crime policy in the other community, which may have led to a higher arrest rate in the comparison group?

^e Several matching methods exist, especially in the context of quantitative program evaluations, to reduce such problems. Propensity Score Matching develops a score to establish the degree of similarity for a potential match based on how well the observable characteristics of individuals in a potential comparison group matches members of the treatment group.² Certain research designs, such as regression discontinuity, utilize program design features to define a similar comparison group. For example, if eligibility for a program is determined by a cut-off (e.g.,

program, if the broader economic conditions differ on average for participants in an employment training program compared to those in its comparison group, it can be difficult to conclude that the difference in outcomes is due to the program versus differences due to the broader economic conditions. A particular challenge for cross-sectional designs is self-selection: even if two participants have the same observed characteristics (age, sex, education level, etc.), if enrolment in a program is voluntary the fact that one chose to participate and the other did not suggests that there are unobserved differences (e.g., one is more motivated) that could cause differences in outcomes regardless of program participation (e.g., a participant may be more motivated than a non-participant and so would have searched harder for a job even if they had not participated in the employment program).

e. *Designs that Utilize both Cross-sectional and Time-series Variation*

It is possible to design an evaluation to utilize both before-after and cross-sectional comparisons. Such a design compares the before-after change in the program participants to the before-after change for those in the comparison

group.^f Because it utilizes both variation over time and variation across the two groups, such a

Cross-Sectional & Time-Series Variation

Building on the previous box, now imagine that the evaluators had good baseline data on arrests in both communities. This design compares the change among program participants against the change for similar individuals in the comparison community. In this case, imagine that the baseline rate of arrests was slightly higher for the comparison group (a reflection of the “tough-on-crime policy) and that these rates had been stable for the period before the program begun. The evaluation reveals that the program was a success: arrest rate fell among program participants but not in the comparison group.

Is the reduction in recidivism the result of the BUILD UP program or the tough-on-crime policy?

Based on the comparison of the change in rates between the two communities before and after the program, we can now say with confidence that the program did have an effect, even though the comparison group lived in a community with a tough-on-crime policy throughout. This highlights how the cross-sectional-times-series design can overcome some problems of cross-sectional alone or time-series alone.

eligibility based on age or income level), regression discontinuity compares participants just below the cut-off against non-eligible individuals with values just above the cut-off, who presumably should be quite similar.

^f Two basic types of such design are longitudinal studies, which follow the same individuals over time for each of the program and comparison groups, i.e., one has multiple observations for each individual corresponding to the

design offers much greater potential to identify a causal relationship between program participation and the achievement of specified outcomes. Although this design can suffer from sources of bias present for both before-after and cross-sectional studies, having two sources of variation can enable one to mitigate the impact of such sources of bias. Evaluators have identified a number of conditions under which it is possible to establish causation using such designs for observational data. While more powerful, such a design requires a good comparison group to those in the program and good baseline data for both groups.

f. *Contribution Analysis*

As noted, in the absence of true randomization to program and non-program settings, one is often left not fully confident that an evaluation has established causality. In such cases, evaluators often look to supplementary analyses, other data, and situational factors to undertake a form of triangulation to establish greater confidence. If multiple, and diverse additional sources of insight (each of which is in and of itself insufficient to establish causation) are all consistent with a causal explanation, then one can have greater confidence in the main study finding. Diversity of sources and approaches is important in this triangulation exercise: if based on the same underlying strategy, even multiple additional types of evidence provide little new information.

Some of these concepts have been formalized under an approach called contribution analysis. Contribution analysis is structured theoretical approach that uses the theory of change (e.g., critical analysis of the causal chain, assumptions, and risks) combined with other sources of evidence (e.g., survey, field notes, administrative data, etc.) from which a reasonable person could conclude that the program produced the outcomes.^{58,59} The limitations of contribution analysis (proclivity to subjective bias,⁵⁸ resource intensive nature,⁵⁸ lack of quality criteria⁶⁰) mean that in and of itself it is not sufficient to assess causality^{58,60,61} for an outcomes-based payment model such as Outcomes Canada, but it can be a useful supplement to other, core evaluation methods.

3.2.5 *Mixed Methods*

Mixed-methods approaches combine in complementary ways different methodological approaches in a single evaluation, seeking to gain the advantages of each method and thereby strengthen the overall study design. Mixed-methods approaches are becoming

before and after periods; and cross-sectional/time-series designs in which the specific individuals in the before-after periods differ within each group. As an example of the latter, imagine a community-based program aimed at increasing knowledge of sexually transmitted diseases among youth aged 16-18. One could survey youth aged 16-20 in two communities before the program is launched in one of them, and then conduct a second, post-program, survey in the two communities three years after implementation. In each community, the specific youth in the before survey would differ from those in the post-program survey. The technique of comparing the difference in the within-group change across the two groups is commonly referred to as “difference-in-difference”.

Mixed Methods

The use of mixed methods requires a mix of outcomes that can be measured through quantitative or qualitative indicators.

Consider an evaluation of BUILD UP Saskatoon where the evaluator not only wants to know if the program worked (quantitative analysis) but why the program worked (qualitative analysis).

Effectiveness (Quantitative) analysis: the evaluator compares the number (or proportion) of participants who had no encounters with the justice system relative to a comparison group over the designated period. To assess this, a survey could ask: *how many interactions have you had with police over the last 6 months?*

Descriptive (Qualitative) analysis: To supplement this information, the evaluator then asks participants these follow-up questions: Q1: Has this program been a good experience for you? Why or why not? Q2: When you think about yourself before entering the program (your thoughts and feelings), do you feel differently about yourself now and if so, how?

commonplace.^{19,20,21,24,41,42,62,63,64,65,66} As noted above in 3.2.2, the term “mixed methods” is commonly used to designate studies that combine both quantitative and qualitative methods. However, we use the term more generally: yes, qualitative and quantitative methods, but also descriptive approaches and effectiveness approaches that seek to establish causality. The effectiveness component (often quantitative) can establish causality (did the program work and to what extent?) while the descriptive (often qualitative component) can provide insight into the underlying mechanisms that contribute to a program’s success (failure) and can provide richer information regarding the impact of a program on the community, program participants, and other relevant stakeholders.

Mixed-methods approaches may be particularly important for the Outcomes Canada platform. Outcomes Canada emphasizes community-led, community-engaged processes to identify promising programs, shape the definition of program outcomes and evaluation design, and to give voice to community organizations and community members, especially program participants. These goals are best

served by descriptive, community-engaged, often participatory approaches. At the same time, establishing whether a program achieves the designated levels of outcomes is central to Outcomes Canada’s outcomes-based payment model of social finance. This goal is best served by effectiveness approaches designed to establish causal connections between a program and the outcomes measured.

Having said this, it must be emphasized that for some settings and designs these two goals will, in the following sense, be in tension. Community-engaged, participatory approaches involve evaluators deeply engaged with program participants (e.g., often seeing program participants

as co-knowledge creators on equal footing with the evaluators). This risks the evaluators becoming part of the intervention—an additional program component specific to the evaluation context absent when a program is delivered more broadly. Such effects are not of great concern for participatory designs because such designs do not seek to make causal inferences regarding effectiveness. They have other goals. However, from the perspective of establishing causal connections between a program and the outcomes observed, these effects “contaminate” the evaluation since they preclude establishing the impact of the program itself, absent evaluator effects. These tensions will require very careful consideration of how the two designs may interact in the evaluation in ways counterproductive to the goal of gaining the advantages of each but instead compromising the strengths of each. Guarding against this may require differential timing of the different components, applying the different approaches to different subsets of study participants, or modifying the approaches in ways that limit the negative interactions while ensuring the core evaluation objectives can still be met.

Short of full, community-engaged participatory approaches, descriptive information regarding program implementation, community members reached, broader community impacts and other similar aspects of a program operation can provide useful information regarding why a program did or did not work (especially in relation to the underlying theory of change) and for successful programs, how to improve future program delivery, challenges to be faced when scaling up, and likely implementation issues in other jurisdictions.^{63,64}

3.2.6 Evaluation Methods Summary

- i. *No one single method is the universally right (or wrong) approach to evaluation but rather the best approach chosen will depend critically on the evaluation objectives — “what does the evaluation seek to demonstrate?”*
- ii. *Evaluations that seek primarily to establish strong causal evidence regarding the impact of a program will draw on methods for establishing causation such as quasi-experimental designs that utilize both cross-sectional and time-series variation when possible. Evaluation that seeks to understand the mechanisms that underlie the outcomes’ causal pathway or participants’ lived experience will prefer community-engaged, possibly participatory, approaches.*
- iii. *Each approach has strengths and weaknesses which evaluators will have to weigh. Causal approaches aim for objectivity and minimizing sources of bias, thus normally limiting the extent to which they seek input from a range of stakeholders including community members or program participants. Evaluations that seek to engage more fully with a range of stakeholders, risk—from the perspective of establishing effectiveness—inducing stakeholder bias that can impact the evaluation’s reliability, validity and importantly transferability. However, such approaches can advance other evaluation goals, especially in the context of Outcomes Canada’s community-focused goals.*

- iv. *Mixed-methods approaches have the potential to marry elements of community focus—a key objective of the Outcomes Canada Platform—to an evaluation of whether a program achieves outcomes as specified in the investment contract. The effectiveness component can establish causation as would be done conventionally while the descriptive, qualitative component can engage more fully with stakeholders including the community and program participants. Such mixed-methods approaches, however, can be challenging and are resource intensive.*

3.3 Methods for Assessing Value for Money

Some contracts will specify outcomes beyond effectiveness alone to include evidence of value for money. More generally, stakeholders, and particularly investors and buyers, may want to know whether the program provides good value for money. Effectiveness is a necessary condition for a program to provide good value for money,⁸ but determining value for money adds considerable complexity to an evaluation and makes substantially greater data demands. An assessment of value for money compares costs and outcomes and therefore requires collecting information on the costs of providing a program. Many methods exist for determining value for money. An important distinction among them is whether they assign a value to the outcomes, and if so, whether that value is expressed in non-monetary or monetary terms.

3.3.1 Approaches that Do not Assign Value to Outcomes

Cost-effectiveness Analysis (CEA) compares the incremental costs of providing a program against the incremental change in outcomes expressed in natural units (e.g., arrests, number of people employed full time, tons of material recycled, number of emergency room visits avoided).⁶⁷ The result is a measure that indicates the additional cost required to obtain an additional unit of the outcome—an indicator of value for money provided by the program. If this number is small (large), then it means that the program produces the outcome at low (high) cost. Because it avoids placing a value on the outcomes but rather leaves them expressed in natural units, it is less complex than the methods that do and is therefore one of the most widely applied approaches. However, because cost-effectiveness can be determined only for each outcome separately, it is best suited for evaluations with a single, or a single dominant, outcome of interest; it is poorly suited for situations in which there are multiple outcomes of interest as it cannot provide an indication of the value for money of the overall program.

⁸ If a program is not effective, there is no reason to proceed further. An ineffective program that does not work can never provide good value for money. Sometimes a program may work but be less effective than existing practice. If it is also sufficiently less costly, such a less effective program may be preferred based on value for money.

3.3.2 Approaches that Do not Assign Value to Outcomes

The most common approaches to assessing value for money by assigning non-monetary value to outcomes have been developed for programs for which the primary outcome is improved health. Consider two health programs, each of which increase the expected year of life for participants by 10 years, but in one case those additional years are lived in full health while in the second people experience severe chronic pain and cannot carry out all their usual activities. Both programs produce the same outcome measured in life-years, but the second produces fewer “quality-adjusted life-years” since individuals live with chronic pain and face limitations on their activity. Cost-Utility Analysis (CUA) assigns a utility weight to each unit (e.g., life year) of the outcome achieved to calculate an outcome measured in quality-adjusted life-years (QALY) produced. With this outcome in hand, which reflects the value to individuals of the health gains produced, Cost-Utility Analysis assesses the value for money of a program by calculating the cost-per QALY produced by the program. Cost-utility analysis has two main advantages over CEA in health contexts: (a) the outcome incorporates consideration of quality of life; and (b) one can assess the value for money of a program with multiple health outcomes, all of which can be expressed in terms of QALYs. A main disadvantage is that CUA can typically only be applied to programs for which health outcomes are the primary outcomes of interest. However, due to a growing interest in non-monetary approaches to measure multi-dimensional aspects of well-being for evaluation, organizations are developing QALY-type measures for other types of outcomes.^{68,69,70}

3.3.3 Approaches that Assign Monetary Value to Outcomes

Diverse approaches exist to assess value for money by assigning monetary values to the outcomes achieved, and thereby provide an indicator of the return on investment (ROI). The dominant approach within program evaluation is cost-benefit analysis (CBA), but before discussing CBA, we want to note a particular type of financial analysis often used to calculate return on investment: cost savings to key stakeholders.

Financial Analysis of Cost Savings focuses solely on the net savings—the difference between the cost of providing a program and the costs averted by reduced service usage by participants—as a measure of

Financial Analysis of Cost Saving

Imagine, the Saskatchewan Ministry for Immigration and Career Training has agreed to fund BUILD UP because the program provides career training to otherwise difficult-to-service groups. An evaluation shows that BUILD UP reduced recidivism, but a financial cost saving analysis reveals no associated cost savings for this Ministry or the provincial government more broadly. Instead, the cost savings accrued solely to local police services.

What might happen to this otherwise successful program if the provincial government’s funding support was based primarily on expected cost savings to the provincial ministry?

return on value. This outcome is understandably of interest to a funder, such as a government ministry, and understanding such financial effects can help identify challenges for broad implementation (e.g., a program may reduce overall costs, but the ministry that provides the program experiences an increase in costs while the costs savings accrue to another ministry). But purely financial analyses provide a very narrow indicator of value for money. Such financial analyses value a program outcome only to the extent that it generates cost savings to a provider/funder; it assigns a value of zero to the effects an outcome has for the individual, their family, or to society more broadly. As such, it is not consistent with the broader principle of program evaluation to consider all effects of a program in society.

Cost-benefit analysis avoids this narrow calculation of only financial impacts by assigning a monetary value to all outcomes of interest and then comparing the program costs against these benefits to assess value for money. The major challenge for cost-benefit analysis is assigning monetary value to outcomes achieved. The preferred economic approach for doing so is to determine a person's willingness-to-pay for the outcome. Cost-benefit practitioners have devised multiple methods for doing this, but they can be difficult and expensive to implement and can raise ethical issues in certain evaluation contexts for which willingness-to-pay is not a good indicator of social value. In some cases, one can use as a proxy a

Cost-Benefit Analysis

The primary objective of SALSA program is youth employment, but the secondary objective is to fund community gardens in communities with little access to fresh fruits and vegetables. SALSA is therefore expected to achieve several financial and non-financial benefits over time that will accrue to different members of society such as the organization, participants, the government, and society more generally.

Financial benefits, e.g.,

- For the SALSA, program profits from salsa sales
- For participants, higher wages.

Non-financial benefits, e.g.,

- For participants of the community garden, health and quality of life gains
- For the community, quality of life gains from the beautification of the neighbourhood and greater diversity of birds, bees and other wildlife.

Financial Costs, e.g.,

- Salsa Ingredients
- Wages
- Hydro
- Property taxes
- Rents

Non-Financial Costs, e.g.,

- Odour from the Salsa production facility
- Increased neighbourhood traffic near garden plots

measure such as the increment to a person’s life-time earnings; this is a relevant proxy for the value of better employment obtained as a result of a training program but also has well-understood biases and limitations. A CBA can in principle include any type of outcome and therefore is the preferred approach when a program generates multiple distinct outcomes of value measured in different natural units (e.g., improved, health, better employment, reduced incarceration). The result of a cost-benefit analysis is expressed by a measure of net benefit (benefits – costs) and represents the net benefit to society of a program across all outcomes of value.

3.3.4 Social Return on Investment (SROI)

SROI attempts to straddle the line between the broader social benefit indicators of success important to a mission-oriented community organization and value for dollar analysis that is of importance to some funders. Essentially, SROI is a cost-benefit analysis (CBA) but with greater emphasis on stakeholder engagement^{41,71,72} and a more in-depth report that should include other social indicators of success beyond just the financial ones. Stakeholders can include beneficiaries, administrators, funders and others as appropriate^{41,51} and, although they may contribute to the evaluation, these stakeholders do not lead the evaluation.⁵¹

However, the judgements of the various stakeholders regarding the weights attributed to external costs/benefits, choice of indicators and other assumptions in the evaluation model can lead to multiple sources of bias.^{51,71} As a result, SROI ratios are not generally comparable across programs.^{10,72} Furthermore, comprehensive SROI analysis is resource intensive in terms of time and specialized skills making it an impractical choice for smaller organizations.⁴¹

3.4 Outcomes and Indicators

Figure 2 depicts the interconnectedness of typical outcomes that community organizations hope to achieve. Whether seeking to address poverty, social justice or inequality, there are multiple pathways to achieving these aspirational goals that are not mutually exclusive. For example, it may not be possible to achieve employment outcomes without also addressing deficits in housing, health or education; and it may not be possible to address educational outcomes without also addressing housing and health and community-level outcomes like safe communities. It is therefore not uncommon to see community organizations such as QUINT Development (BUILD UP Saskatoon) and Spence Neighbourhood Association (SALSA) pursuing

Figure 2: Interconnectedness of Outcomes



multiple outcomes through multiple programs at any given time. Within specific programs, indicators for these outcomes, whether for employment, education, health, housing, community and environment, are the concrete measures used to represent the outcome to assess short, medium and longer-term outcomes.

Outcomes are generally defined at the conceptual level (e.g., reduced unemployment), but for evaluation purposes, they need to be translated into operational indicators capable of measuring the outcome. Outcomes are operationalized through empirical, data-based indicators. Often, a given outcome can be measured in a number of ways. For example, employment can be measured in terms of presence or absence of a job, number of hours worked, whether employed part time or full time, whether wage is subsidized or not subsidized, hourly rate or wage level, which is often as a proxy for quality of employment. The process of moving from conceptual outcomes to operational indicators requires that the outcomes be specified sufficiently and precisely to be able to be measured meaningfully by indicators. Vague or aspirational outcomes (e.g., improve opportunities for youth in the Spence neighbourhood) need to be unpacked and translated into very specific indicators of change (e.g., proportion of youth graduating from high school and entering college). Selecting the appropriate indicators can be complex requiring certain assumptions, particularly when outcomes are not easily observable (e.g., optimistic about the future – a proxy for mental health).

The choice of indicator to represent an outcome is one of the most important choices in an evaluation. No matter how carefully designed using the most sophisticated methods, an indicator that poorly represents the true objective will not provide useful information. Very carefully measuring the wrong thing is not useful. There can be a tension between a preferred research design based purely on considerations of design (e.g., ability to establish causation) but which requires use of a less-preferred indicator that is a less good proxy for the underlying outcome sought versus use of a less-preferred design that will enable one to use a preferred outcome indicator.

Table 1 lists some of the common outcomes and corresponding indicators for the types of programs expected to be of interest to the Outcomes Canada Platform. The table is not meant to be exhaustive but rather provides examples of outcomes and indicators, and the linkage between outputs and their corresponding outcomes, as well as highlight potential data sources for those outcomes and indicators. The specific indicator and best source of data will depend on the program of interest but, over time, this table can be supplemented as the platform develops. In many cases, obtaining outcomes data specific to program recipients will require primary data collection but, in some cases, outcomes data will be available in administrative databases if it is possible to obtain access (e.g., arrest records for program participants; cost of

Table 1: Common Outcomes

The outcomes table below is not an exhaustive list of outcomes and indicators but rather provides examples that may be of interest to an outcomes-based payment model such as Outcomes Canada. Output indicators illustrate the difference between these output indicators and their corresponding outcome indicators (see the text for a fuller description of the differences). Note that any given outcome can be operationalized in a number of ways depending on the goals of the program and the evaluation. Finally, the last column links outcome indicators to potential data sources at the national, provincial and local level.

Common Outcomes	Output Indicators	Outcome Indicators	Outcome Data Sources (National, Provincial/Territorial, Local)
Employment	Job Readiness and Skills Training (# people completed job search training, job interview training, resume building, reading, numeracy, writing, high school equivalent, acquiring ID, ESL, attended a job interview)	Job Type (# any job, part-time/full-time, temporary, precarious, subsidized/unsubsidized)	National Only: <ul style="list-style-type: none"> ● Statistics Canada Socioeconomic Time Series Data — formerly known as CANSIM (e.g., permanent employees, temporary employees) National and Provincial/Territorial: <ul style="list-style-type: none"> ● Statistics Canada Socioeconomic Time Series Data (e.g., employed full-time or part-time) ● Statistics Canada Labour Force Survey (e.g., job permanency, union status, reason for leaving job). Data available also for selected Census Metropolitan Areas (e.g., Toronto, Winnipeg). National, Provincial/Territorial and Local:

Common Outcomes	Output Indicators	Outcome Indicators	Outcome Data Sources (National, Provincial/Territorial, Local)
			<ul style="list-style-type: none"> ● Statistics Canada Labour Force Survey (with special permission)
	Workshops (<i># complete, # attendees</i>)	Hours Worked (<i>per day, week, month</i>)	<p>National Only:</p> <ul style="list-style-type: none"> ● Statistics Canada Socioeconomic Time Series Data <p>National and Provincial/Territorial:</p> <ul style="list-style-type: none"> ● Statistics Canada Labour Force Survey (also for selected Census Metropolitan Areas e.g., Toronto, Winnipeg) <p>National, Provincial/Territorial and Local:</p> <ul style="list-style-type: none"> ● Statistics Canada Labour Force Survey (with special permission)
	On-the-job Training (<i># internships/apprenticeships completed, hours in training, # with new skills certification</i>)	Quality of Work (<i>hourly wage, change in wage, receipt of benefits, type of benefits</i>)	<p>National Only:</p> <ul style="list-style-type: none"> ● Statistics Canada Socioeconomic Time Series Data <p>National and Provincial/Territorial:</p> <ul style="list-style-type: none"> ● Labour Force Survey (also for selected Census Metropolitan Areas e.g., Toronto, Winnipeg)

Common Outcomes	Output Indicators	Outcome Indicators	Outcome Data Sources (National, Provincial/Territorial, Local)
			<ul style="list-style-type: none"> ● Statistics Canada Canadian Community Health Survey <p>National, Provincial/Territorial and Local:</p> <ul style="list-style-type: none"> ● Statistics Canada Labour Force Survey (with special permission, includes a much richer dataset) ● Statistics Canada Canadian Community Health Survey (with special permission)
		Time to Employment <i>(weeks, months)</i>	<p>National and Provincial/Territorial:</p> <ul style="list-style-type: none"> ● Statistics Canada Socioeconomic Time Series Data <p>National, Provincial/Territorial and Local:</p> <ul style="list-style-type: none"> ● Statistics Canada Labour Force Survey (with special permission)
		Long-term Employment <i>(never unemployed, unemployed days, unemployed)</i>	<p>National and Provincial/Territorial:</p> <ul style="list-style-type: none"> ● Statistics Canada Labour Force Survey (also for selected Census Metropolitan Areas e.g., Toronto, Winnipeg)

Common Outcomes	Output Indicators	Outcome Indicators	Outcome Data Sources (National, Provincial/Territorial, Local)
			<ul style="list-style-type: none"> ● Statistics Canada Canadian Community Health Survey <p>National, Provincial/Territorial and Local:</p> <ul style="list-style-type: none"> ● Statistics Canada Labour Force Survey (with special permission) ● Statistics Canada Canadian Community Health Survey (with special permission)
		<p><i>Cost Savings (social assistance, housing, other social services)</i></p>	<p>National and Provincial/Territorial:</p> <ul style="list-style-type: none"> ● Statistics Canada Labour Force Survey (also for selected Census Metropolitan Areas e.g., Toronto, Winnipeg) ● Statistics Canada Canadian Community Health Survey <p>National, Provincial/Territorial and Local:</p> <ul style="list-style-type: none"> ● Statistics Canada Labour Force Survey (with special permission) ● Statistics Canada Canadian Community Health Survey (with special permission) <p>Province/Territory-specific data, example:</p>

Common Outcomes	Output Indicators	Outcome Indicators	Outcome Data Sources (National, Provincial/Territorial, Local)
			<ul style="list-style-type: none"> New Brunswick Institute for Research, Data and Training (NB-IDRT)
Justice	(# internship, # in job training, # in rehabilitation)	Recidivism (<i>any incarceration, long-term incarceration, arrests, court appearances</i>)	<p>Publicly available data generally only provide information on the number of contacts with the justice system (e.g., incarcerations, crimes, arrests) without information on the number of contacts had by the same person.</p> <p>National and Provincial/Territorial:</p> <ul style="list-style-type: none"> Statistics Canada Corrections Key Indicator Report for Adults and Youth Statistics Canada Youth Custody and Community Services, Integrated Correctional Services Survey and Canadian Correctional Services Survey
		Cost savings (<i>police, justice</i>)	<p>National and Provincial/Territorial:</p> <ul style="list-style-type: none"> Statistics Canada Adult Correctional Services Survey (operating expenditures and average daily cost per inmate) <p>Province/Territory-specific data, example:</p>

Common Outcomes	Output Indicators	Outcome Indicators	Outcome Data Sources (National, Provincial/Territorial, Local)
			<ul style="list-style-type: none"> • Ellingwood/Public Safety Canada, Research report 2015-R018 (average police cost per year by office type)
Homelessness	Housing units built, temporary shelter provided (by type of shelter), long term housing provided.	Long-term housing <i>(any housing, public housing, subsidized housing, low-income housing, independent living, required supports)</i>	National and Provincial/Territorial: <ul style="list-style-type: none"> • Statistics Canada Canadian Housing Survey (also for selected Census Metropolitan Areas) National, Provincial/Territorial, Sub-provincial/-territorial: <ul style="list-style-type: none"> • Statistics Canada Census of the Population Local only: <ul style="list-style-type: none"> • Canadian Observatory on Homelessness • Local municipal and community organizations estimating the number of homeless in their area (e.g., City of Hamilton, Ontario, Point in Time Connection data)
	Social service supports <i>(identification, numeracy, budgeting, other)</i>	Time housed <i>(specified period, permanently housed, return to street)</i>	National and Provincial/Territorial: <ul style="list-style-type: none"> • Statistics Canada Canadian Housing Survey (also for selected Census Metropolitan Areas)

Common Outcomes	Output Indicators	Outcome Indicators	Outcome Data Sources (National, Provincial/Territorial, Local)
		Cost Savings (<i>public housing, subsidized units, social services</i>)	Local only, examples providing isolated data points: <ul style="list-style-type: none"> ● Auditor General of Ontario, Annual Report 2017. Chapter 3, section 3.14 “Social and Affordable Housing” ● Mental Health Commission of Canada, “National Final Report – Cross-Site At Home/Chez Soi Project” 2017. ● Jadidzadeh, Falvo and Dutton, “Cost Savings of Housing First in a Non-Experimental Setting”, Canadian Public Policy, 2020 Vol 46(1): 22-36.
Health (mental and physical)	Food security (<i># meals served, # food bank clients, # vegetables consumed</i>)	Reliance on substance use (<i>any alcohol, drugs, tobacco, # alcohol, drugs, tobacco, # of alcohol, drugs, tobacco free days</i>)	National: <ul style="list-style-type: none"> ● Statistics Canada Canadian Tobacco and Nicotine Survey (CTNS) National and Provincial/Territorial: <ul style="list-style-type: none"> ● Statistics Canada Canadian Community Health Survey (CCHS) ● Government of Canada Public Health Infobase Public Health Agency of Canada

Common Outcomes	Output Indicators	Outcome Indicators	Outcome Data Sources (National, Provincial/Territorial, Local)
			<ul style="list-style-type: none"> ● Canadian Student Tobacco, Alcohol and Drugs Survey (CSTADS) (provinces only, students in grades 7-12) <p>National, Provincial/Territorial, and Local:</p> <ul style="list-style-type: none"> ● CTNS and CCHS (with special permission) <p>Province/Territory-specific data, examples:</p> <ul style="list-style-type: none"> ● Ontario only: Public Health Ontario Interactive Opioid Tool ● Manitoba only: Manitoba Population Research Data Repository <p>Local:</p> <ul style="list-style-type: none"> ● Simcoe Muskoka Opioid Overdose Cohort (SMOOC)
	Health Care Access (<i># physician visits, # dental visits, # mental health visits</i>)	Health Metrics (<i>blood pressure and others, self-reported health</i>)	<p>National and Provincial/Territorial:</p> <ul style="list-style-type: none"> ● Statistics Canada Canadian Community Health Survey (CCHS) ● Government of Canada Public Health Infobase Public Health Agency of Canada

Common Outcomes	Output Indicators	Outcome Indicators	Outcome Data Sources (National, Provincial/Territorial, Local)
			<ul style="list-style-type: none"> ● Statistics Canada Canadian Student Tobacco, Alcohol and Drugs Survey (CSTADS): provinces only, students in grades 7-12 National, Provincial/Territorial, and Local: <ul style="list-style-type: none"> ● CCHS (with special permission)
	Provision of Basic Needs <i>(# accessed housing, # accessed food bank, # hygiene products distributed)</i>	Improved Quality of Life (QoL) <i>(increased social participation, self-esteem, self-sufficiency, self-reported QoL, Activities of Daily Living, family reunification, reduced feelings of exclusion and discrimination)</i>	National and Provincial/Territorial: <ul style="list-style-type: none"> ● Statistics Canada Canadian Community Health Survey (CCHS) National, Provincial/Territorial, and Local: <ul style="list-style-type: none"> ● CCHS (with special permission)
	Outreach (# of new clients)	Cost Savings <i>(emergency services, hospital days, other social services)</i>	National and Provincial/Territorial: <ul style="list-style-type: none"> ● Canadian Centre on Substance Use and Addiction: Canadian Substance Use Costs and Harms (CSUCH) data
Economic Development	Business Development (# of business loans or	Business Development (# of new businesses opening,	National, Provincial/Territorial, and Local:

Common Outcomes	Output Indicators	Outcome Indicators	Outcome Data Sources (National, Provincial/Territorial, Local)
	<i>grants, # entrepreneurial workshops, # participant in mentorship program)</i>	<i># businesses recording a profit, # businesses with X employees, # of educated youth returning to the community)</i>	<ul style="list-style-type: none"> ● Statistics Canada Business Register
Environmental	Recycling Education (<i>workshops, media activities, educational materials circulated</i>)	Recycled Materials (<i>Kg diverted from landfill</i>)	<p>National Only:</p> <ul style="list-style-type: none"> ● Statistics Canada Households and the Environment Survey ● Statistics Canada Survey of Innovation and Business Strategy <p>National and Provincial/Territorial:</p> <ul style="list-style-type: none"> ● Statistics Canada Waste Management Industry Survey <p>National, Provincial/Territorial and Local:</p> <ul style="list-style-type: none"> ● Statistics Canada Households and the Environment Survey (with special permission) <p>Local only:</p> <ul style="list-style-type: none"> ● Municipal sources of data e.g., City of Toronto Solid Waste & Diversion Rates

Common Outcomes	Output Indicators	Outcome Indicators	Outcome Data Sources (National, Provincial/Territorial, Local)
	Recycling Activities (<i># items recycled, lbs of recycled material \$ value of items recycled</i>)	Reduction in Energy Use (<i>kilowatt hours, water litres, gigajoules gas</i>)	National and Provincial/Territorial: <ul style="list-style-type: none"> ● Government of Canada, Canada Energy Regulator Local only: <ul style="list-style-type: none"> ● Provincial or sub-provincial sources (e.g., BC Hydro Power smart, Vancouver Island)
	Community Activities (<i>Tree planting, green box program, blue box program</i>)	Air Quality (<i>CO2 levels</i>)	Local: <ul style="list-style-type: none"> ● Government of Canada, National Air Pollution Surveillance (NAPS) Program, data for over 700 collection stations in Canada
	Energy Retrofits (<i>commercial, residential</i>)	Cost Savings (<i>public buildings, home heating bills, home hydro bills, home water bills</i>)	National: <ul style="list-style-type: none"> ● Statistics Canada Households and the Environment Survey: Energy Use National, Provincial/Territorial and Local: <ul style="list-style-type: none"> ● Statistics Canada Households and the Environment Survey: Energy Use* (with special permission)
	Energy Audits (<i>commercial, residential, public buildings</i>)		National and some Provinces:

Common Outcomes	Output Indicators	Outcome Indicators	Outcome Data Sources (National, Provincial/Territorial, Local)
			<ul style="list-style-type: none"> ● Government of Canada, Natural Resources Canada, Survey of Household Energy Use

Indicators

The two primary outcomes for BUILD UP Saskatoon are reduced contact with the justice system and reduced justice costs. Secondary outcomes include consistent employment and family reconciliation.

Quantifiable indicators: Number or proportion of participants who had any interaction with the justice system.

Non-Quantifiable indicators: Participants are asked in a survey about their relationship with their family. Specifically, whether their relationship is better or worse than it was before.

Objective Indicators: Administrative records that document number and type of interactions participants had with the system.

Subjective Indicators: In a survey, using an ordinal scale from 1-5, participants rate their level of satisfaction with the BUILD UP program.

Self-Reported Indicators: Participants self-declare their interactions with the justice system.

Third-party Reported Indicators: Parole officer reports participant has had no interaction with the justice system.

Proxy: Because evaluators could not access justice data due to privacy concerns, employed full-time is used as a proxy for not incarcerated. The rationale is if a participant is employed full-time, it is unlikely that they are incarcerated.

providing a service). More generally, even when such data cannot be linked to program participants, these data can provide important context and background regarding community-level rates of outcomes of interest. More on indicators and data below.

3.4.1 Types of Indicators

The appropriate indicator(s) for any given outcome will be dependent on what the evaluation wants to demonstrate as well as the evaluation approach best suited to facilitate the answers taking into consideration constraints (resource, time and data) and various contextual matters.

Indicators should be specific, identifiable, reliable and appropriate such that they measure what they are intended to measure and are fit for purpose. In situations where it is very difficult or impossible to represent an outcome, rough indicators serve merely as proxies of the outcome recognizing that they are imperfect.

Some of the key characteristics of indicators are as follows.

a. *Quantifiable*

Quantifiable indicators are a numerical representation of an outcome such as absolute count (1,2,3...), scale or index (1-5), ratio or percentage (%).

b. *Non-Quantifiable*

Non-quantifiable indicators include verbal expressions of an outcome of interest, written documents, and

other non-numeric outcome indicators. Even where it is possible to develop quantitative indicators for a construct (e.g., job satisfaction can be rated on a scale from 1 to 5 with 1 representing “not at all satisfied” and 5 representing “very satisfied”) non-numeric data convey additional information.

c. *Objective*

Objective indicators are quantifiable, transparent and do not depend primarily on individual judgement. For example, the absolute number of program participants employed after one year.

d. *Subjective*

Subjective indicators are judgement-based. They are the opinion or interpretation of a participant or an observer. For example, one’s happiness (happy, neither happy nor unhappy, unhappy).

e. *Self-Reported*

Self-reported indicators can be either objective or subjective. For example, self-reported health status (good, neither good nor poor, poor) or weight recorded in a diary (pounds per week).

f. *Third-Party Reported*

Third party reported indicators are observations about the outcome of interest made by someone other than the participant. For example, a participant’s physician may report the participant’s blood pressure at regular intervals.

3.5 Data Types and Accessibility

Data access, quality and consistency can be significant barriers to rigorous evaluation and assessment.^{3,8,9,10,11,73,74} The type of data available and accessible will influence both the evaluation approach and methods used. Additionally, the quality and consistency of the data can make some data sources less (more) reliable than others. Evaluators will often face trade-offs between resource limitations, data quality limitation and data availability and access.

3.5.1 Primary Data

Primary data are collected first-hand by the evaluation team using methods such as surveys, interviews, focus groups or observation. The depth and breadth of the informational demands of the evaluation design will dictate to some degree the data collection method. For example, a survey is a versatile tool for collecting quantitative data (e.g., number employed after 6 months) and, to a more limited degree, qualitative data (e.g., in what way did the program improve your employability?). Focus groups and interviews are well suited for collecting stakeholders’ viewpoints and lived experience (e.g., how has the program improved your feelings about the future?). Diary and sensors are a good option for capturing individual-level data at regular intervals (e.g., minutes of daily exercise). However, primary data are only as good as the collection methods and level of participant compliance. Not all participants will

consistently make journal entries, answer any or all questions on a survey, use a sensor as prescribed, etc. Primary data collection tends to be time and resources intensive. To ensure data quality and consistency, primary data collection requires continual monitoring.^{73,74} Embedding evaluation activities such as data collection into the program design can improve efficiency³⁹ reducing the burden of these activities on organizations and other stakeholders including the program recipients.⁷⁴

3.5.2 Secondary Data

Secondary data are collected by a third party and include: survey data collected by agencies for which data collection is their core mission such as census and survey data collected by Statistics Canada, administrative data such as that collected in the process of administering a program or service (e.g., arrest records by a law enforcement agency, recipients of subsidized housing by a social housing agency), and, increasingly, third-party data collected through a variety of digital programs, services, and activities (e.g., tweets from twitter accounts, population movements through phone tracking technologies). In some cases, such secondary data are the preferred source of data because it provides the most accurate values (e.g., unemployment claims can provide more accurate data on receipt of unemployment benefits than self-report), but, in others cases, such data provide proxy measures at best because they were collected to serve purposes other than what the evaluation seeks to measure (e.g., neighbourhood income as a proxy for individual-level income). Secondary data can also suffer from issues of data quality and consistency, making verification of the quality essential. Secondary data can be less expensive and less resource intensive to obtain and use, though this is not always the case. Certain types of secondary data— especially identifiable data pertaining to specific individuals—can be difficult and time-consuming to access, requiring ethics agreements and data request forms and other processes.

Secondary Data

BUILD UP Saskatoon has been preparing for a formal evaluation of this program over the past year but accessing data has proven challenging. For costing data, BUILD UP has reach out to legal aid, the Ministry of Corrections, Policing and Public Safety and other relevant contacts but after several months of on-going communication, the evaluation team was informed by these groups that they do not know how or where to locate this information. The team has also been in contact with police services for both costing information and remand and dispatch records, but progress has been slow. Negotiations are on-going.

4. Did the program achieve the specified thresholds for the outcomes?

The objective of an outcomes-based payment evaluation is to determine whether program outcomes reached the agreed-upon thresholds to trigger payment. At one level this is straightforward, especially if the thresholds have been clearly specified in the same terms that the evaluation measures the outcomes (which a good design should do). In some cases, a highly successful program will have exceeded a threshold unequivocally, and in others an ineffective program will have unequivocally not done so. In many other cases, however, the answer is not straightforward. If the contract specifies multiple outcomes, some thresholds may have been achieved and others not. One solution in such a situation is to pro-rate the overall payment to reflect partial success, which requires agreeing on the weight attached to each outcome since they will not all be of equal importance. In other cases, the estimates will be borderline. For instance, the point estimate of the impact on an outcome could exceed the threshold but the statistical confidence interval could overlap the threshold, so the estimate is not statistically different from the threshold. Unless this possibility was anticipated when specifying the threshold, deciding how to handle this may require consultation and negotiation among stakeholders. Such situations can best be avoided by negotiating clear, unambiguous, measurable outcome thresholds when negotiating the investment contract, and ensuring that the design of the evaluation will directly answer the question of whether a threshold has been attained.

VI. Conclusion

With limited public dollars available to support community-based social and environmental programs, community organizations need new and innovative ways to engage with private investors. The Outcomes Canada model aims to address the challenge by creating a multi-sector initiative through which community-led programs will produce positive, measurable outcomes, such as economic well-being, skills training and job creation, improved health, or other social and environmental outcomes. A critical element of Outcomes Canada's outcomes-based payment model of social financing is outcomes measurement and evaluation.

The roadmap identifies critical components of an evaluation, and key issues associated with each component, and options to consider in addressing each question. Because each evaluation must be customized to each evaluation setting, the roadmap cannot recommend specific methods in general. Furthermore, the roadmap reflects the Outcomes Canada's ambitious model that incorporates community-grounded, community-led approaches to program development with a goal of rigorous evaluation of program outcomes. These twin aims will at times come into tension, a tension that Outcomes Canada will have to manage. Community-engaged evaluation approaches seek to give voice to community members regarding their lived experience. The close relationship between the participants and the evaluators means that, at

times, the evaluation can effectively become part of the intervention. In contrast, more classical evaluation approaches that seek to establish whether a program caused certain outcomes emphasize objectivity and a certain kind of distance between evaluators and a program. One approach is not better than the other; they simply have different goals. Attaining both requires mixed-methods evaluation approaches using innovative designs that address these tensions to avoid compromising both aspects. Such mixed-methods approaches will be challenging and resource intensive.

The roadmap is a first step for measurement and evaluation to support the Outcomes Canada platform. Outcomes Canada will need to develop its measurement and evaluation approaches more comprehensively in a way that reflects how it will approach the measurement and evaluation work of the platform. To aid this work, it will be important for Outcomes Canada to define more clearly the types of programs it seeks to support and develop, and especially the kinds of community contexts, the relationships it envisions, the way it will conceive outcomes, the expectations of community organizations and their communities and of investors, along with many other issues. Each community program supported through Outcomes Canada will not be completely distinct—they will have unique elements but will also share many characteristics. From this, it may be possible to identify a few basic models of measurement and evaluation best suited for the anticipated set of priority contexts, enabling Outcomes Canada to match each program to a basic measurement and evaluation model that can be refined to accommodate the unique elements of a given program. As part of this work, Outcomes Canada will have to make important decisions about how much and what types of measurement and evaluation infrastructure it wants to build in-house and what it wants to contract with external parties. This includes both people infrastructure—staff with required measurement and evaluation expertise—and physical infrastructure to, for instance, store and analyze data gathered through evaluation, administer surveys, and other essential activities.

References:

1. Canadian Task Force on Social Finance(2010) “Mobilizing Private Capital for Public Good”. *Finance*, (December), p. 40. Available at: https://mccconnellfoundation.ca/wp-content/uploads/2017/07/FinalReport_MobilizingPrivateCapitalforPublicGood_30Nov10.pdf.
2. Community Foundations of Canada(2020) “Outcomes Canada Model Overview”.
3. Lalande, L. and Cave, J.(2017) “Measuring Outcomes in Practice: Fostering an Enabling Environment for Measurement in Canada. Mowat Research #157”. Available at: <https://munkschool.utoronto.ca/mowatcentre/measuring-outcomes-in-practice/>.
4. Reisman, J. and Olazabal, V.(2016) “Situating the Next Generation of Impact Measurement and Evaluation for Impact Investing”The Rockefeller Foundation. Available at: <https://assets.rockefellerfoundation.org/app/uploads/20161207192251/Impact-Measurement-Landscape-Paper-Dec-2016.pdf> (Accessed: 10 February 2020).
5. Sillanpää, V.(2013) “Measuring the Impacts of Welfare Service Innovations”. *International Journal of Productivity and Performance Management*, 62(5), p. 474–489. doi: 10.1108/IJPPM-08-2012-0094.
6. Cukier, W. and Gagnon, S.(2017) “Social Innovation: Shaping Canada’s Future”. Available at: <https://www.ryerson.ca/content/dam/diversity/reports/social-innovation-shaping-canadas-future.pdf> (Accessed: 5 February 2020).
7. Rawhouser, H. *et al.*(2019) “Social Impact Measurement: Current Approaches and Future Directions for Social Entrepreneurship Research”. *Entrepreneurship Theory and Practice*, 43(1), p. 82–115. doi: 10.1177/1042258717727718.
8. Impact Measurement Project(2019) “Using Self-reported Data for Impact Measurement: How to Use Stakeholder Surveys to Improve Impact Performance”. Available at: <https://impactmanagementproject.com/stakeholder/using-self-reported-data-for-impact-measurement/>.
9. Barraket, J. and Yousefpour, N.(2013) “Evaluation and Social Impact Measurement Amongst Small to Medium Social Enterprises: Process, Purpose and Value”. *Australian Journal of Public Administration*, 72(4), p. 447–458. doi: 10.1111/1467-8500.12042.
10. Nicholls, J. *et al.*(2012) “A Guide to Social Return on Investment”. Available at: http://www.socialvalueuk.org/app/uploads/2016/03/The_Guide_to_Social_Return_on_Investment_2015.pdf (Accessed: 6 January 2021).
11. Clark, C. and Brennan, L.(2016) “Social Entrepreneurship: A Global Model for Evaluating

- Long-Term Impact”. *International Journal of Entrepreneurship*, 20, p. 1–15. Available at: <https://www.ryerson.ca/content/dam/diversity/reports/social-innovation-shaping-canadas-future.pdf>.
12. Ruff, K. and Olsen, S.(2018) “The Need for Analysts in Social Impact Measurement”. *American Journal of Evaluation*, 39(3), p. 402–407. doi: 10.1177/1098214018778809.
13. European Commission(2014) “Proposed Approaches to Social Impact Measurement”. Available at: https://ccednet-rcdec.ca/sites/ccednet-rcdec.ca/files/proposed_approaches_to_social_impact_measurement.pdf.
14. Hadad, S. and Gauca, O. (Drumea)(2014) “Social Impact Measurement in Social Entrepreneurial Organizations”. *Management & Marketing*, 9(2), p. 119–136. Available at: <http://libaccess.mcmaster.ca/login?url=https://search.proquest.com/docview/1618069639?accountid=12347>.
15. Moore, M.-L. *et al.*(2012) “Social Finance Intermediaries and Social Innovation”. *Journal of social entrepreneurship*, 3(2), p. 184–205. Available at: https://uwaterloo.ca/waterloo-institute-for-social-innovation-and-resilience/sites/ca.waterloo-institute-for-social-innovation-and-resilience/files/uploads/files/social_finance_intermediaries.pdf.
16. Trelstad, B.(2016) “Making Sense of the Many Kinds of Impact Investing”. *Harvard Business Review*. Available at: <https://hbr.org/2016/01/making-sense-of-the-many-kinds-of-impact-investing> (Accessed: 2 November 2021).
17. Rockefeller Centre(n.d.) “Impact Investing: An Introduction”. Available at: https://www.rockpa.org/wp-content/uploads/2017/10/RPA_PRM_Impact_Investing_Intro_WEB.pdf (Accessed: 2 November 2021).
18. Canadian Evaluation Society(2015) “What is Evaluation?” Available at: <https://evaluationcanada.ca/what-is-evaluation> (Accessed: 20 May 2021).
19. Simuyemba, M. C. *et al.*(2020) “Real-time Evaluation Pros and Cons: Lessons from the Gavi Full Country Evaluation in Zambia”. *Evaluation: The International Journal of Theory, Research and Practice*, 26(3), p. 367–379. doi: <http://dx.doi.org/10.1177/1356389019901314>.
20. Walmsley, B.(2012) “Towards a Balanced Scorecard: A Critical Analysis of the Culture and Sport Evidence (CASE) Programme”. *Cultural Trends*, 21(4), p. 325–334. doi: <http://dx.doi.org/10.1080/09548963.2012.726800>.
21. Wholey, J. S. *et al.*(2010) ‘*Handbook of Practical Program Evaluation*. 3rd edn’(Wiley Online Books).

22. Pauly, B. *et al.* (2014) "Approaches to Evaluation of Homelessness Interventions". *Housing, Care and Support*, 17(4), p. 177–187. doi: 10.1108/HCS-07-2014-0017.
23. Carleton University (2020) "Common Approach to Impact Measurement". Available at: <https://carleton.ca/commonapproach/key-elements/>.
24. Centers for Disease Control and Prevention (CDC) (2011) "Developing an Effective Evaluation Plan". Available at: <https://www.cdc.gov/obesity/downloads/cdc-evaluation-workbook-508.pdf> (Accessed: 24 February 2021).
25. FRITZ, J. (2020) "How to Talk About Nonprofit Impact from Inputs to Outcomes". *The Balance Small Business*. Available at: <https://www.thebalancesmb.com/inputs-outputs-outcomes-impact-what-s-the-difference-2502227> (Accessed: 16 September 2021).
26. HM Treasury (2018) "The Green Book: Central Government Guidance on Appraisal and Evaluation" GOV.UK. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/685903/The_Green_Book.pdf (Accessed: 20 November 2018).
27. McEwen, N. (2002) "Towards Evidence-based Policy for Canadian Education". in Broucker, P. de and Sweetman, A. (eds) *Towards Evidence-based Policy for Canadian Education*. Montreal and Kingston: McGill Queen's University Press, p. 61. Available at: <http://libaccess.mcmaster.ca/login?url=https://www.proquest.com/books/towards-evidence-based-policy-canadian-education/docview/59926429/se-2?accountid=12347>.
28. Centers for Disease Control and Prevention (CDC) (2007) "Types of Evaluations". Available at: [https://www.cdc.gov/std/Program/pupestd/Types of Evaluation.pdf](https://www.cdc.gov/std/Program/pupestd/Types%20of%20Evaluation.pdf) (Accessed: 3 March 2021).
29. Centers for Disease Control and Prevention (2008) *'Introduction to Process Evaluation in Tobacco Use Prevention and Control'* Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. Available at: <http://www.cdc.gov/tobacco/publications/index.htm>. (Accessed: 6 May 2021).
30. Gertler, P. J. *et al.* (2016) *'Impact Evaluation in Practice, Second Edition'* Washington, DC: Inter-American Development Bank and World Bank. doi: doi:10.1596/978-1-4648-0779-4.
31. Shahidur R. Khandker *et al.* (2009) *'Handbook on Impact Evaluation: Quantitative Methods and Practices'* World Bank Publications.
32. Rogers, P. (2020) "Real-Time Evaluation. Monitoring and Evaluation for Adaptive Management Working Paper Series, Number 4". Available at:

<https://www.betterevaluation.org/en/monitoring-and-evaluation-adaptive-management-working-paper-series> (Accessed: 19 July 2021).

33. McNall, M. and Foster-Fishman, P. G.(2007) “Methods of Rapid Evaluation, Assessment, and Appraisal”. *American Journal of Evaluation*, 28(2), p. 151–168. doi: 10.1177/1098214007300895.
34. Williams, E.(2021) “Rapid Evaluation”BetterEvaluation. Available at: <https://www.betterevaluation.org/en/approach/rapid-evaluation> (Accessed: 19 July 2021).
35. Intrac for Civil Society(2017) “Real-Time Evaluation”. Available at: <https://www.intrac.org/wpcms/wp-content/uploads/2017/01/Real-time-evaluation.pdf>.
36. USAID(2010) “Performance Monitoring and Evaluation: Tips Using Rapid Appraisal Methods”. Available at: <https://web.archive.org/web/20130302141625/http://transition.usaid.gov/policy/evalweb/documents/TIPS-UsingRapidAppraisalMethods.pdf> (Accessed: 19 July 2021).
37. Woodsong, N. *et al.*(2005) “Qualitative Research Methods: A Data Collector’s Field Guide”. Available at: https://pdf.usaid.gov/pdf_docs/PNADK310.pdf (Accessed: 27 July 2021).
38. Watkins, D. C.(2017) “Rapid and Rigorous Qualitative Data Analysis: The “RADaR” Technique for Applied Research”. *International Journal of Qualitative Methods*, 16(1), p. 1609406917712131. doi: 10.1177/1609406917712131.
39. Canadian Homelessness Research Network(2013) “What Works and For Whom? A Framework for Designing and Implementing Promising Practices Research - Part 2”Toronto. Available at: https://www.homelesshub.ca/sites/default/files/attachments/PPFramework_Part2.pdf (Accessed: 2 September 2020).
40. Mercier, C. *et al.*(1992) “Program Evaluation of Services for the Homeless: Challenges and Strategies”. *Evaluation and Program Planning*, 15(4), p. 417–426. Available at: <https://www.sciencedirect.com/science/article/abs/pii/014971899290079A>.
41. Arvidson, M. *et al.*(2013) “Valuing the social? The Nature and Controversies of Measuring Social Return on Investment (SROI)”. *Voluntary Sector Review*, 4(1), p. 3–18. doi: <http://dx.doi.org/10.1332/204080513X661554>.
42. Nastasi, B. K. and Hitchcock, J.(2009) “Challenges of Evaluating Multilevel Interventions”. *American Journal of Community Psychology*, 43(3–4), p. 360–376. doi: <http://dx.doi.org/10.1007/s10464-009-9239-7>.

43. Meschede, T. and Chaganti, S.(2015) “Home For Now: A Mixed-methods Evaluation of a Short-term Housing Support Program for Homeless Families”. *Evaluation and Program Planning*, 52, p. 85–95. doi: 10.1016/j.evalprogplan.2015.03.009.
44. Sette, C.(2021) “Participatory Evaluation” BetterEvaluation. Available at: https://www.betterevaluation.org/en/plan/approach/participatory_evaluation (Accessed: 20 July 2021).
45. Guijt, I.(2014) “Participatory Approaches” Florance. Available at: https://www.betterevaluation.org/sites/default/files/Participatory_Approaches_ENG_0.pdf.
46. Zukoski, A. and Luluquisen, M.(2002) “Participatory Evaluation What is it? Why do it? What are the challenges? Policy Brief (5)”. Available at: <https://www.betterevaluation.org/sites/default/files/Evaluation.pdf> (Accessed: 20 July 2021).
47. Andrews, N. C. Z. *et al.*(2019) “Research and Evaluation with Community-Based Projects: Approaches, Considerations, and Strategies”. *American Journal of Evaluation*, 40(4), p. 548–561. doi: <http://dx.doi.org/10.1177/1098214019835821>.
48. USAID(2011) “Performance Monitoring and Evaluation Tips: Conducting a Participatory Evaluation”. Available at: https://pdf.usaid.gov/pdf_docs/pnadw101.pdf (Accessed: 19 July 2021).
49. Community Tool Box(2021) “Participatory Evaluation”. in *Evaluating Community Programs and Initiatives* Center for Community Health and Development at the University of Kansas. Available at: <https://ctb.ku.edu/en/table-of-contents/evaluate/evaluation/participatory-evaluation/main> (Accessed: 20 July 2021).
50. Intrac for Civil Society(2017) “Participatory Evaluation”. Available at: <https://www.intrac.org/wpcms/wp-content/uploads/2017/01/Participatory-evaluation.pdf> (Accessed: 20 July 2021).
51. Social Value UK(2016) “Supplementary Guidance on Stakeholder Involvement”. Available at: [https://www.socialvalueuk.org/app/uploads/2016/03/Supplementary_Guidance_on_Stakeholder_Involvement_\(PDF\).pdf](https://www.socialvalueuk.org/app/uploads/2016/03/Supplementary_Guidance_on_Stakeholder_Involvement_(PDF).pdf) (Accessed: 17 March 2021).
52. Deaton, A. and Cartwright, N.(2018) “Understanding and Misunderstanding Randomized Controlled Trials”. *Social Science & Medicine*, 210, p. 2–21. doi: <https://doi.org/10.1016/j.socscimed.2017.12.005>.
53. Hanley, P. *et al.*(2016) “Reassessing RCTs as the “Gold Standard”: Synergy Not Separatism in Evaluation Designs”. *International Journal of Research & Method in Education*, 39(3), p. 287–298. doi: 10.1080/1743727X.2016.1138457.

54. Gibson, B. and Flaherty, D.(2017) “Employment Impact Assessments: A Review of Methodologies” International Labour Organization. Available at: <http://libaccess.mcmaster.ca/login?url=https://search.proquest.com/docview/2063163818?accountid=12347>.
55. Johnston, D.(2006) “Lessons To Be Learnt? The Role of Evaluations of Active Labour Market Programmes in Evidence-based Policy Making”. *Public administration and development*, 26(4), p. 329–338. doi: <http://dx.doi.org/10.1002/pad.405>.
56. Hujer, R. and Caliendo, M.(2000) “Evaluation of Active Labour Market Policy: Methodological Concepts and Empirical Estimates”. *IDEAS Working Paper Series from RePEcSt. Louis: Federal Reserve Bank of St Louis*. Available at: <http://libaccess.mcmaster.ca/login?url=https://search.proquest.com/docview/1698640410?accountid=12347>.
57. Smith, J. and Sweetman, A.(2016) “Viewpoint: Estimating the Causal Effects of Policies and Programs”. *Canadian Journal of Economics/Revue canadienne d'économique*, 49(3), p. 871–905. doi: <https://doi.org/10.1111/caje.12217>.
58. Budhwani, S. and McDavid, J.(2017) “Contribution Analysis: Theoretical and Practical Challenges and Prospects for Evaluators”. *Canadian Journal of Program Evaluation*, 32. doi: [10.3138/cjpe.31121](https://doi.org/10.3138/cjpe.31121).
59. Mayne, J.(2008) “Contribution Analysis: An Approach to Exploring Cause and Effect”. Available at: https://www.researchgate.net/publication/46472564_Contribution_analysis_An_approach_to_exploring_cause_and_effect (Accessed: 1 November 2021).
60. Bannister, J. and O’Sullivan, A.(2013) “Knowledge Mobilisation and The Civic Academy: The Nature of Evidence, The Roles of Narrative and The Potential of Contribution Analysis”. *Contemporary Social Science*, 8(3), p. 249–262. doi: [10.1080/21582041.2012.751497](https://doi.org/10.1080/21582041.2012.751497).
61. Noltze, M. *et al.*(2014) “Contribution Analysis as an Evaluation Strategy in the Context of a Sector-Wide Approach: Performance-Based Health Financing in Rwanda”. *African Evaluation Journal; Vol 2, No 1 (2014)*. doi: [10.4102/aej.v2i1.81](https://doi.org/10.4102/aej.v2i1.81).
62. Burns, S. and Cupitt, S.(2003) “Managing Outcomes: A Guide for Homelessness Organisations”.
63. Perezniето, P. and Taylor, G.(2014) ‘Review of Evaluation Approaches and Methods Used By Interventions on Women and Girls’ Economic Empowerment’. *Overseas Development Institute, Mar 2014, 62 pp*. Overseas Development Institute. Available at: <http://libaccess.mcmaster.ca/login?url=https://search.proquest.com/docview/1559001244?accountid=12347>

countid=12347.

64. Bambra, C. *et al.* (2005) "Does "Welfare-To-Work" Work? A Systematic Review of the Effectiveness of the UK's Welfare-To-Work Programmes for People With A Disability or Chronic Illness". *Social Science & Medicine*, 60(9), p. 1905–1918. doi: 10.1016/j.socscimed.2004.09.002.

65. Housing and Homeless Partnership(n.d.) "Halifax Housing First Evaluation Framework".

66. Calgary Homeless Foundation(2017) "System Planning Framework". Available at: http://calgaryhomeless.com/content/uploads/SSPF_V116_2017-03-15.pdf (Accessed: 11 September 2020).

67. Drummond, M. *et al.* (2015) 'Methods for the Economic Evaluation of Health Care Programmes, 4th ed'. *Methods for the Economic Evaluation of Health Care Programmes*.

68. Engel, L.(2017) 'Going Beyond Health-Related Quality of Life for Outcome Measurement in Economic Evaluation' Simon Frazer University.

69. Al-Janabi, H. *et al.* (2012) "Development of a Self-Report Measure of Capability Wellbeing for Adults: The ICECAP-A". *Quality of Life Research*, 21(1), p. 167–176. doi: 10.1007/s11136-011-9927-2.

70. Crawford, L. *et al.* (2013) "Does Education Need a QALY and Is LAYS It?" *Centre for Global Development*. Available at: <https://www.cgdev.org/blog/does-education-need-qaly-and-lays-it> (Accessed: 1 October 2021).

71. Cordes, J. J. (2017) "Using Cost-Benefit Analysis and Social Return On Investment to Evaluate the Impact of Social Enterprise: Promises, Implementation, and Limitations". *Evaluation and Program Planning*, 64, p. 98. Available at: <http://libaccess.mcmaster.ca/login?url=https://www.proquest.com/docview/1944229634?accountid=12347>.

72. Pathak, P. and Dattani, P. (2014) "Social Return On Investment: Three Technical Challenges". *Social Enterprise Journal*, 10(2), p. 91–104. doi: <http://dx.doi.org/10.1108/SEJ-06-2012-0019>.

73. Turner, A. (2015) "Performance Management in a Housing First Context: A Guide for Community Entities" Toronto. Available at: <https://www.homelesshub.ca/sites/default/files/attachments/CEGuide-final.pdf> (Accessed: 9 September 2020).

74. Wallace, B. *et al.* (2015) "Shifting the Evaluative Gaze: Community-based Program Evaluation in the Homeless Sector". *Gateways: International Journal of Community Research*

and Engagement, 8(1), p. 43–58. doi: 10.5130/ijcre.v8i1.3936.

75. Ramakers, A. *et al.* (2017) “Not Just Any Job Will Do: A Study on Employment Characteristics and Recidivism Risks After Release”. *International journal of offender therapy and comparative criminology*. 2016/07/28, 61(16), p. 1795–1818. doi: 10.1177/0306624X16636141.