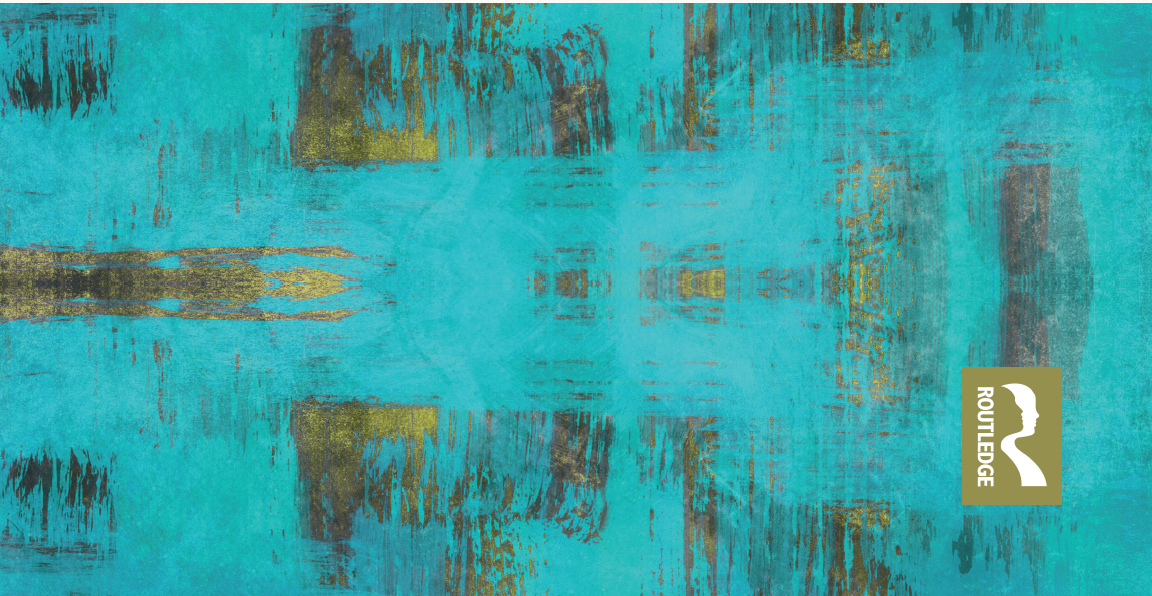
Comparative Policy Evaluation

POLICY EVALUATION IN THE ERA OF COVID-19

Edited by Pearl Eliadis, Indran A. Naidoo, and Ray C. Rist



"The diversity and richness of the topics show immediately how important the book is. The findings … to name a few, [such as] the deprioritization of evaluation during the crisis, the taken-for-grantedness of a number of underlying assumptions, describes [the] many efforts of evaluators."

Frans L Leeuw, *Professor Law, Public Policy and Social Sciences, Maastricht University*

"*The strength of the book is the combination of the perspectives and the relevant insights they present. The book is undoubtedly timely, relevant, and important, not only for understanding what happened during the past years, but also in preparing for what will come.*"

Per Øyvind Bastøe, *Associate professor, VID Specialized University, Oslo, Norway*

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Policy Evaluation in the Era of COVID-19

Did evaluation meet the challenges of the COVID-19 crisis? How were evaluation practices, architectures, and values affected? *Policy Evaluation in the Era of COVID-19* is the first to offer a broad canvas that explores government responses and ideas to tackle the challenges that evaluation practice faces in preparing for the next global crisis. Practitioners and established academic experts in the field of policy evaluation present a sophisticated synthesis of institutional, national, and disciplinary perspectives, with insights drawn from developments in Australia, Canada, and the UK, as well as the UN.

Contributors examine the impacts of evaluation on socioeconomic recovery planning, government innovations in pivoting internal operations to address the crisis, and the role of parliamentary and audit institutions during the pandemic. Chapters also examine the Sustainable Development Goals, and the inadequacy of human rights-based approaches in evaluation, while examining the imperative proposed by some authors that it is time that we take seriously the call for substantial transformation.

Written in a clear and accessible style, *Policy Evaluation in the Era of COVID-19* offers a much-needed insight on the role evaluation played during this unique and critical juncture in history.

Pearl Eliadis is Associate Professor (Professional) at the Max Bell School of Public Policy at McGill University in Montreal, Quebec. She also has a private law practice with a focus on human rights, national institutions, and democratic development and has worked with the United Nations (UN), the European Union (EU), and the Organisation for Economic Co-operation and Development (OECD) in several sub-Saharan and Asian countries on a range of national and global projects at the intersection of evaluation and the rule of law. She is a member of the Quebec Bar Association and the Ontario Law Society and held the position of president of the Human Rights section of the Quebec Bar for three terms until 2021. She also lectures at McGill's Faculty of Law and is a full member of McGill's Centre for Human Rights and Legal Pluralism.

Indran A. Naidoo is the Director of the Independent Office of Evaluation of the International Fund for Agricultural Development (IFAD) Rome, Italy. As former Director of the Independent Evaluation Office of the United Nations Development Programme (UNDP), New York, he transformed the evaluation function in terms of independence, output, and engagement, producing over 150 country and corporate evaluations. The National Evaluation Capacity series grew to become the largest and most influential platform for advancing evaluation at the government level. Previous roles included that of Deputy Director-General of M&E at the independent Public Service Commission and M&E Director at the Department of Land Affairs. He served as Vice-Chair of UN Evaluation Group (UNEG), instructor at the International Program for Development Evaluation Training (IPDET), and was a founding board member of the South African Monitoring and Evaluation Association (SAMEA). He holds a DPhil and post-graduate degree in Geography and Education and is the recipient of numerous awards and extensively published.

Ray C. Rist is one of the creators and co-directors of the original IPDET and former president of the International Development Evaluation Association (IDEAS). Retired from the World Bank, Dr. Rist continues to advise organizations and national governments throughout the world on how to design and build results-based M&E systems. His career includes 15 years in the U.S. Government with senior appointments in both the executive and legislative branches. He has held professorships at John Hopkins, Cornell, and George Washington Universities and has been a Fulbright Fellow at the Max Planck Institute in Berlin. He has authored and edited 35 books, written more than 140 articles, and lectured in more than 85 countries.

Comparative Policy Evaluation
Edited by Ray C. Rist

The Comparative Policy Evaluation series is an interdisciplinary and internationally focused set of books that embodies within it a strong emphasis on comparative analyses of governance issues—drawing from all continents and many different nation states. The lens through which these policy initiatives are viewed and reviewed is that of evaluation. These evaluation assessments are done mainly from the perspectives of sociology, anthropology, economics, policy science, auditing, law, and human rights. The books also provide a strong longitudinal perspective on the evolution of the policy issues being analyzed.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Policy Evaluation in the Era of COVID-19

Edited by
Pearl Eliadis, Indran A. Naidoo, and Ray C. Rist

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

First published 2023
by Routledge
605 Third Avenue, New York, NY 10158

and by Routledge
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

Library of Congress Cataloging-in-Publication Data
Names: Eliadis, F. Pearl, 1959- editor. | Naidoo, Indran A., editor. | Rist, Ray C., editor.
Title: Policy evaluation in the era of Covid-19 / Edited by Pearl Eliadis, Indran A. Naidoo, and Ray C. Rist.
Description: New York, NY : Routledge, 2023. | Series: Comparative policy evaluation | Includes bibliographical references and index. |
Identifiers: LCCN 2022049291 (print) | LCCN 2022049292 (ebook) | ISBN 9781032452968 (hardback) | ISBN 9781032452975 (paperback) | ISBN 9781003376316 (ebook)
Subjects: LCSH: Political planning--Evaluation--Cross-cultural studies. | COVID-19 Pandemic, 2020--Cross-cultural studies. | Sustainable development--Evaluation--Cross-cultural studies. | Human rights--Evaluation--Cross-cultural studies.
Classification: LCC JF1525.P6 P652 2023 (print) | LCC JF1525.P6 (ebook) | DDC 303.48/57072--dc23/eng/20230105
LC record available at https://lccn.loc.gov/2022049291
LC ebook record available at https://lccn.loc.gov/2022049292

ISBN: 978-1-032-45296-8 (hbk)
ISBN: 978-1-032-45297-5 (pbk)
ISBN: 978-1-003-37631-6 (ebk)

DOI: 10.4324/9781003376316

Typeset in Bembo Std
by KnowledgeWorks Global Ltd.

Contents

List of Contributors

Maria Barrados is currently the Executive-in-Residence at the Sprott School of Business and serves as Adjunct Professor, Department of Sociology and Anthropology, both at Carleton University. She is the former president of the Public Service Commission of Canada.

Jim Blain graduated from the University of London (UK) in economics before joining the Canadian federal public service where he held various positions in program evaluation. He has retired from this position as Director General, Program Evaluation, Human Resources and Skills Development, Canada, and has continued to consult government on evaluation issues for a number of years.

Pearl Eliadis is Associate Professor (Professional) at the Max Bell School of Public Policy at McGill University. She is also a lawyer with an international law practice in human rights, national institutions, and democratic governance. She has worked for organizations such as the UN and OECD on a range of issues at the intersection of evaluation and the rule of law.

Jan-Eric Furubo has worked at different Swedish ministries, governmental agencies, and The National Audit Office. He has published works about the role of evaluation in democratic societies and was co-editor of the *International Atlas of Evaluation* (2002), *Evaluation – Seeking Truth or Power?* (2011), *Evaluation in Turbulent Times* (2013), and *The Evaluation Enterprise* (2020). He was the first president of the Swedish Evaluation Society and has had different roles within the European Evaluation Society.

Robert Lahey, President of REL Solutions Inc., was founding head of the Canadian Government's Centre of Excellence for Evaluation. He has advised globally (World Bank, UN agencies, UNEG, specific countries, international networks) on capacity building of national monitoring and evaluation systems, and evaluation use in supporting implementation of the SDGs.

Jeremy Lonsdale was a Director at the National Audit Office, UK until 2021, latterly being responsible for defense value for money studies. He worked on secondment as a Senior Research Leader at RAND Europe

2014–2016. His publications include as co-editor of "*Performance Auditing: Contributing to Accountability in Democratic Government*" (2011).

Dorothy Lucks is the Executive Director of Sustainable Development Facilitation (SDF) Global, a social enterprise based in Australia, that supports strategic and evidence-based sustainable development processes globally, including active evaluation practice in complex contexts. Dorothy is a Credentialed Evaluator and Fellow of the Australian Evaluation Society.

Steve Montague, partner of Performance Management Network Inc., is a credentialed evaluator, a Fellow of the Canadian Evaluation Society, and has four decades of experience as a practitioner in performance measurement and program evaluation as a management consultant, an adjunct professor, and as an evaluation manager in a major Canadian federal government department.

Indran A. Naidoo is the Director of the Independent Office of Evaluation of the International Fund for Agricultural Development (IFAD). He has served as Director of the Independent Evaluation Office of UNDP and Deputy Director-General of the Office of the Public Service Commission of South Africa. Indran is a globally recognized evaluation thought leader, speaker, manager, and author, who has spearheaded international reform initiatives.

Michael Quinn Patton is Founder and CEO of the Utilization-Focused Evaluation organization, US, and has written several books on evaluation. He sat on the faculty of the University of Minnesota for 18 years and is also a former President of the American Evaluation Association. He continues to be an active trainer and workshop presenter.

Ray Pawson is the Emeritus Professor of Social Research Methodology at the University of Leeds, UK. He is best known for pioneering the "realist" approach to program evaluation. His books include *Realistic Evaluation* (1997), *Evidence-based Policy* (2006), and *The Science of Evaluation* (2013). A new monograph, *How to Think Like a Realist*, is in preparation.

Ray C. Rist was one of the creators and co-directors of IPDET and also a Former President of the International Development Evaluation Association (IDEAS). Retired from the Independent Evaluation Group of the World Bank, Dr. Rist continues to advise organizations and national governments throughout the world on how to design and build results-based M&E systems.

Peter Wilkins is an Adjunct Professor at The John Curtin Institute of Public Policy (JCIPP) at Curtin University. He served as Western Australia's Deputy Ombudsman. Prior to this, he served as Western Australia's Assistant Auditor General Performance Review. He is a National Fellow and Western Australia's Fellow of the Institute of Public Administration Australia.

Acknowledgments

Many thanks to the Independent Office of Evaluation of the International Fund for Agricultural Development (IFAD) for their aid in editing this text.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Introduction[1]

Pearl Eliadis, Indran A. Naidoo, and Ray C. Rist

The COVID-19 pandemic exposed and deepened economic, social, political, and environmental fault lines on a compressed timescale. In developing countries and fragile states in particular, new layers of conflict and vulnerability emerged. Hard-won progress towards reaching the Sustainable Development Goals (SDGs) began to dwindle and dissipate. Democratic accountability measures were limited as governments suspended legislatures in many countries, and ceded power to executive branches of government. Marginalized and vulnerable populations in all countries experienced further and disproportionate adverse impacts because of public measures that were designed to limit the spread of disease for the general population. The resulting social, political, and democratic deficits fostered impunity for human rights violations and reduced accountability.

Governments had to respond quickly. But the speed that was needed to generate viable solutions, as well as the accompanying resource implications, left little room for traditional forms of evaluation. Many evaluators appear to have been caught flat-footed by restrictions resulting from the pandemic environment, especially limits on travel and onsite work. Border closures, travel restrictions, and quarantine limited in-person evaluations and personal contacts with clients, informants, and colleagues. The demand for discrete studies based on simple attribution models declined, while a move towards broader knowledge streams, more capable of responding to current crises, accelerated. Few evaluations were able to address the implications of the new democratic deficits nor the human rights violations that were revealed or exacerbated by emerging forms of inequality: these are not areas of comfort for traditional evaluation practice. Evaluators were being asked to move beyond methodological problems and technical solutions to engage with substantive and high-level inquiries that were content-driven, consistent with human rights norms, and informed by what has been learned during the pandemic.

The task was made more complex as the measures taken by states in response to the crisis differed in many ways from other kinds of

DOI: 10.4324/9781003376316-1

emergencies. The 2007–2009 global financial crisis prompted interventions to address capital flight, market instability, weakened support to overseas development assistance, and to manage contagious stresses in the financial system. What we learned then has relevance to what we know now about rapid and massive interventions in both fiscal and monetary policy. However, this pandemic was the product of a health emergency and had other dimensions. In terms of other pandemics, it is true that diseases such as the 1918 Spanish flu were also emergencies at a global scale. Indeed, the Spanish flu killed about three times more people than COVID-19 (based on figures available at the time of writing). But the devastating impacts of the virus were very much the product of contemporary economic and social factors: interconnected economies and more developed health and social safety nets in the 2020s were able to provide assistance to populations at a much larger scale than at any time in history. As well, institutional and human rights constraints for democracy limited what governments could and could not do in the name of public emergencies (even if not all governments paid heed to those standards). Many of those constraints, which include the international human rights law framework, did not exist before the Second World War.

National and international agencies mobilized on an unprecedented scale to develop effective vaccines and to roll out vaccination programmes. As Ray Pawson writes in his chapter in this volume, the enormity of the COVID-19 response reached down:

> from macro-economic strategies to counteract mushrooming international debt, sweeping onward to propose comprehensive controls on every institution, organization, and service, [ending] in draconian restrictions on all individual behaviour and contact. In short, the policy response under consideration here consisted of an unparalleled exercise in social control and a sociological explanation is required to account for its fragility.

Another important feature of this pandemic was that measures to fight the virus were taken almost synchronously (albeit using different measures) at a global level. This context offered opportunities and challenges for evaluators to assess the impacts and "what worked," and to do so in nearly real time. Communications, messaging, misinformation, and disinformation were accelerated and often distorted by digital and social media, supported by technological advances. For many organizations, oversight and evaluation activities were suspended.

In a world where the inter-related priorities of public health, the environment, human security, democracy, and human rights were in rapid flux, evaluation professionals became engaged in a larger battle about and for truth and evidence in what Michael Quinn Patton describes in his

chapter as a post-truth, anti-science political world: "That is the context we should consider as we extract lessons about transforming evaluation to deal with these challenges."

This book is about whether we can articulate those lessons and whether evaluators provided these insights. Did evaluation meet the challenges of this unique and critical juncture? How were evaluation practice, architecture, and values affected? How will this learning be reflected in future practice? What were evaluators able to say about "what worked" and how did they fare compared to other knowledge producers? How are evaluators going to shift knowledge into a "big tent" while retaining the relevance of evaluation, especially as a programmatic accountability tool?

This volume attempts to offer tentative perspectives rather than answers. Even as pandemic fatigue seemed to be leading governments to wind down measures like lockdowns and masking, countries like the US and South Africa were experiencing new surges of infection in mid-2022. This book is therefore not being written from an "it is over, so what did we learn" angle, but rather from the viewpoint that much of what has happened will continue, change, and become part of a new and very different reality. The nine chapters take several institutional, national, and disciplinary perspectives to explore both the shortcomings of evaluation but also its innovations and successes. Insights are drawn from previous volumes in this series, while examining the imperative proposed by some authors to take seriously the call for substantial transformation.

The first section of this introduction provides an overview of what evaluators were focusing on and writing about at the outset of the pandemic. It is based on a meta-analysis of the literature review that was conducted for this book based on what evaluation organizations were doing and writing about in the first year of the pandemic. The topics addressed begin with the relatively operational and granular considerations that first appeared during the pandemic, and gradually evolved to address concerns about the future of evaluation and its mission. The second section of this introduction builds on some of the themes raised by the meta-analysis. It focuses on the role of evaluation in turbulent times, systems thinking and complexity, the future of the SDGs, and the critical importance of human rights-based approaches.

Meta-analysis: What were evaluators focusing on?

As part of the research undertaken for this book, a meta-analysis was conducted of evaluation publications in English and in Spanish in the first year of the pandemic (March 2020–February 2021). We reviewed the work of 26 evaluation bodies, namely, four multilateral development banks,[2] four United Nations (UN) entities,[3] eight regional evaluation societies,[4] eight country-level evaluation societies,[5] as well as the International

Development Evaluation Association (IDEAS), and the Organisation for Economic Co-operation and Development (OECD; DaC EvalNet)/ United Nations Development Programme (UNDP). Of these, 11 organizations had produced sufficient relevant material to be included in the analysis (Arumburu, 2021).[6] While the meta-analysis did not purport to be comprehensive, it did offer an interesting snapshot of what evaluators were initially focusing on, namely: rescoping and redesigning evaluation practice; ethical considerations and human rights; preparedness for the future; demonstrating evaluation value in contested contexts; and tools and methodologies. Each of these areas will be dealt with in turn.

Rescoping and redesigning evaluation practice

The first category that emerged from the literature review focused on three areas: changes to work planning, real-time evaluations, and collaboration and communications. Adjustments in evaluation scope and processes took place relatively early in the pandemic. Organizations reassessed their priorities in light of new limitations, as well as the risks of expediting processes. Some organizations issued updated guidelines and guidance, including joint guidance, bearing in mind that the SDGs and the 2030 Agenda were rapidly being revisited (Independent Evaluation Office/United Nations Development Programme & Organisation for Economic Co-operation and Development/ Development Assistance Committee [IEO/UNDP & OECD/DAC], 2020). Strategies were applied throughout the evaluation cycle: for example, feasibility assessments were used to adjust evaluation scope and assess project criticality (Office of Evaluation and Oversight, n.d.). There were also examples of mid-cycle adjustments: the World Bank's Independent Evaluation Group examined rating processes and methodologies to account for shocks and to implement mid-course corrections, providing more room to help projects recover and meet targets later (Independent Evaluation Group, 2020).

Another major shift was the demand for real-time evaluation. Advice and recommendations were needed to support planning, responses, and recovery efforts. Development and humanitarian cooperation also shifted to meet the needs of developing countries that were responding to the pandemic. The World Bank underscored the importance of evaluations in crisis settings to inform projects about what works and what does not. It was noted that this approach likely requires a new mindset regarding the independence of evaluation offices because real-time evaluation requires working closely with implementation teams (Independent Evaluation Group, 2020).[7]

On the third point, proactive collaboration and communications were emphasized by the UNDP, the Inter-American Development Bank, and the International Fund for Agricultural Development. In different ways,

these organizations highlighted knowledge-sharing among institutions, networks, and regions as the key to navigating complex crises. The pandemic also offered opportunities to improve the use of information and communications technology (ICT) to support and develop local capacity (Raimondo, Vaessen, & Branco, 2020). Preliminary planning and training for remote evaluations helped to assuage some of these concerns, and to develop or adapt research questions and techniques that could be conducted with local and international teams using a new generation of ICT platforms (Hassnain, 2020). New collaboration platforms emerged during the pandemic, including the COVID-19 Global Evaluation Coalition, a new partnership that focused on improving coordination and national capacity in measurement and evaluation. Importantly, the Global Evaluation Coalition emphasizes synergies and learning, while reducing duplication in evaluating different elements of COVID-19 pandemic responses (Global Evaluation Coalition, n.d.).

Tools and methodologies

The meta-analysis also emphasized the measurement of the evaluative impact of both COVID-19 and the actions taken to address the pandemic. In this regard, many evaluations had to rely on desk-based assessments. Evaluators were clearly aware of the risks of remote analysis for both scope and depth, as well as the risks of bias and exclusion (Buchanan-Smith, 2021).

Evaluators also reached for technology and remote data sources, including geospatial data and artificial intelligence. Social media and big data have also been prominent in discussions about the future of evaluation. Although there is little documented evidence about ICT tools and their effectiveness in evaluation, they obviously provided critical communications solutions during the pandemic. Techniques include remote/online smartphone-based surveys, online focus groups, social media, participatory processes mediated by video, and geo-spatial technologies from satellite or aerial imagery.

As for big data and artificial intelligence, these sources of information have had a growing role in policy because results can be delivered in close to real time (Petersson & Breul, 2017). But evaluation is not as familiar with big data and geo-spatial analysis tools, according to Indran A. Naidoo in his chapter. This gap was evident in the Socio-Economic Response and Recovery Plans (SERPs) based on reviews from non-traditional evaluation sectors such as research and academic think-tanks. It also became apparent during the pandemic that we know very little about most of the algorithms behind big data or about the impacts of artificial intelligence such as facial recognition technologies, and what we do know is often troubling (Office of the Privacy Commissioner of Canada, 2022).

⋆⋆⋆

In conclusion, the meta-analysis suggested that that the immediate concerns of evaluators were highly pragmatic in nature, but beyond these important practical problems and challenges, it was clear that evaluators were feeling the pressure to answer larger questions, consolidate knowledge streams and work more collaboratively with other disciplines and professionals, and at the same time to engage in critical inquiry about ethical frameworks.

Ethics and human rights

There is little consensus on ethics for evaluators, and perspectives tend to be fragmented and inconsistent. Ethics are a contested concept, with viewpoints ranging from a purely economic, neo-liberal view to a wide range of sociological and environmental approaches (van den Berg, 2022). Having said that, it is not yet clear whether ethics in evaluation changed as a result of the pandemic. What does seem clear is that certain elements assumed more importance in contexts of lower public accountability and transparency, along with higher risk, fragility, conflict, and impunity (Office of Internal Oversight Services [OIOS], 2020). These included "do no harm," prioritizing safety, proactive communication, considering biases, and ensuring inclusivity. Others, such as attention to legal standards such as human rights, were deficient even in re-stated ethical frameworks (Eliadis, this volume).

One area of concern was risk of bias, noting the exclusionary potential of ICT for those populations who were already difficult to reach before the pandemic. In particular, the principles of "do no harm" and inclusivity were seen as critical considerations for assessing whether evaluations should proceed at all during the pandemic. Reduced access to the field and reliance on remote techniques increased the need for time and effort to plan and conduct evaluations strategically and to ensure that the evaluations were genuinely insightful and could confer meaningful benefits in a period of stressed resources and enhanced risk (OIOS, 2020).

Transparency is another key element: in times of crisis, it required decision-makers and evaluators to engage with massive resource mobilizations and budget reallocations, especially where corruption was a substantial risk. Audit offices were expected to track spending to provide legislators with transparency about how these enormous resource mobilizations were spent and managed under emergency circumstances. The role of state audit offices and of internal audit and evaluation services has proved to be critical during this period according to chapters by Maria Barrados and Jeremy Lonsdale, and by Maria Barrados, Steve Montague, and Jim Blain.

Re-examining evaluation ethics and principles as a result of the crisis may not come from traditional programme review, but rather from harder questions about central ethical concerns in an emergency context:

- Impartiality: Is impartiality still possible in a pandemic? From limited capacity to engage with hard-to-reach populations to reliance on big data or artificial intelligence that may reinforce bias, impartiality is an ongoing concern. People may experience limits in accessing ICT and these limits may introduce sample selection bias or reinforce asymmetries in data collection based on factors such as gender and literacy (Center on Gender Equity and Health, UCSD et al.,2021 ; Independent Office of Evaluation, 2021).
- Disclosure: Alternatives to in-person methodologies require a wide range of sampling or interview techniques, such as key informant interviews, but the use of such means to circumvent methodological restrictions raises another ethical issue in terms of the disclosure of methodological constraints and ensuring transparency (Buchanan-Smith, 2021).
- "Do no harm": Through the literature, this ethical principle had renewed relevance, given the impact of emergency responses on vulnerable groups and those directly affected by lockdowns and mobility restrictions. Evaluators must ensure safety, health and welfare, and respect social/physical distancing. While there may be no complete substitute for in-person and onsite evaluations (decisions about who to interview, e.g., are fundamental to the evaluation process), strategies to offset these concerns include more focused engagement with local researchers and evaluators. Nonetheless, in some countries, respondents may be more vulnerable due to national or local controls over the internet and covert surveillance that increase the capacity to geo-locate survey responders (Independent Office of Evaluation, 2021; OIOS, 2020).
- Equity: Although local engagement and training take time, the pandemic enhanced the importance of supporting and developing the capacity and engagement of local researchers and evaluators, thus building diversity and inclusion in the team. A related challenge is that of ensuring the availability of teams with trained female evaluators. Gender imbalances or disproportionate impacts increased the challenges in recruiting and deploying women evaluators, especially in countries with existing social and cultural restrictions (Center on Gender Equity and Health, UCSD et al., 2021).
- Equality and non-discrimination: Linked to the previous point, pre-existing gender inequalities and discriminatory social norms were exacerbated during the pandemic, resulting in inequitable sharing of the disease burden. These include strict social norms that commonly translate into forced confinement, domestic and sexual violence, and limited movement outside the home for many women and girls. Access to basic services – including information, physical and mental health, and other support mechanisms – was restricted. The diversion of already-limited health resources towards the fight against

> the pandemic has strained resources available for pre- and post-natal health care and family planning. Protection services for survivors of gender-based violence, whether government or non-governmental, have notably been impacted by the pandemic. All these factors may render the likelihood of reaching these populations even more remote (Center on Gender Equity and Health, UCSD et al., 2021; United Nations Population Fund, 2020).

Finally, some organizations relied on rights-based evaluations and more equity-focused and gender-responsive interventions. These areas of focus were mainly limited to UN entities (see, e.g., United Nations, 2020). They emphasized the importance of disaggregated data for disadvantaged sub-populations. Evaluation teams were encouraged to be even more deliberate about limitations and biases by diversifying the samples and ensuring that marginalized and vulnerable groups are included. In an analysis of ethical frameworks and standards issued in the pandemic, Pearl Eliadis points out in her chapter in this book that ethical standards and principles for evaluators (especially outside the UN context) offered little in the way of normative human rights content. Guidelines on ethics lacked organizing principles to distinguish among various ethical standards or to decide which standards matter more. Few differentiated between optional norms and those with legal force, let alone those standards that are grounded in the international human rights framework.

Preparedness for the future

New practices and approaches have presented opportunities to improve evaluation practice in the future, including the potential to reduce our collective environmental footprints, strengthen local engagement, and integrate human rights-based approaches with particular sensitivity to systemic/structural forms of discrimination and disadvantage. Relevance will not simply be about doing more and better, it may be about doing less and paying attention to the need to do things differently.

Second, and as previously noted, organizations emphasize the importance of enhancing local capabilities and building national evaluation capacity, especially to support the SDGs (Center on Gender Equity and Health, UCSD et al., 2021; OIOS, 2020). During health crises in particular, remote work is essential, and the use of local consultants and researchers is a priority (Furubo & Stame, 2019). Building local teams, in turn, requires experienced evaluators with established networks in-country who can leverage local capacity.

Preparing evaluation for the "post-normal" future will clearly require organizations to mine the learnings from COVID-19, while incorporating information and lessons about previous shocks. Publications reviewed in the meta-analysis centred on sharing lessons and resources.

Many organizations provided resource libraries and revisited information and guidance about the SDGs. Organizations in this category included the World Bank, the UNDP, the European Evaluation Society, and the Australian Evaluation Society. Keeping the big picture in mind and incorporating different knowledge streams appears to facilitate speaking to the future needs of policy and programme evaluation (Forss et al., 2021).

Demonstrating evaluation's value proposition

Early on, it was apparent that the COVID-19 pandemic would not be just a short-term disruption. The review identified a clear concern among evaluators to ensure their own value proposition by demonstrating evaluation's value in contested contexts. From an analysis conducted by the UN Office of Internal Oversight Services (OIOS) of 11 evaluation guidelines across the UN system and the World Bank, there was a consensus among all the entities that produced COVID-19 evaluation guidelines that it is no longer "business as usual" for evaluation. The evaluation function needed to repurpose and adapt its focus and approach to reflect the limitations posed by the crisis and ensure utility in supporting organizational responses (OIOS, 2020).

A level up from these more operational and cooperation preoccupations is the question of transformation. To address the complexity of the global crises confronting us, from pathogens to environmental crises, authors have called for a transformation of how we approach public policy problems and, more importantly for the purposes of this volume, how we approach evaluation (see, e.g., van den Berg et al., 2021). As well, in his contribution to this volume, Patton proposes three transformative approaches to achieve the kind of sea change that is needed to grapple with the challenges before us: moving from project thinking to systems thinking, moving from theory of change to theory of transformation, and engaging seriously with the implications for evaluation of complexity. Many of these big picture themes form narrative threads for this book and its inquiry into evaluation, not only in times of crisis but also for what lies ahead.

Evaluation in times of crisis

Evaluation, with its focus on accountability and good governance, should have been well-positioned to respond to the crisis. But, as contributor Naidoo points out in his chapter entitled, "Implications for Evaluation, What We Learn from the UN and Country COVID-19 Response Plans, and Reflecting in Future Scenarios," it is far from clear that evaluation systems actually rose to the challenge. The evaluation sector had evolved in environments that possessed a degree of stability; governments had predictable and structured planning processes with clearly established sets of

users for reporting results. Naidoo argues that the COVID-19 crisis disrupted many of these systems and processes, as well as the connections among them; these weaknesses in the evaluation system were revealed by the SERPs.

Discrete interventions have little value in all-of-government or all-of-society approaches. Instead, what is required is ideological and behavioural changes from evaluation which has historically remained disengaged from policy and operational interventions; according to Patton in his chapter, the new context requires more engagement. Naidoo's observations are echoed in the pre-pandemic literature which observes the pressure on evaluation to address uncertainties and risks. As Thomas A. Schwandt observed, in a remarkably prescient article just before the pandemic, that such pressure arose not only within the particular framework of discrete interventions, but also in the design and management of interventions. Schwandt argued that new approaches to evaluation practice linked to planning and decision-making are more likely to reflect assumptions of unpredictability, imperfect information, instability, and pluralism in value determination (Schwandt, 2019).

If Schwandt was right, these approaches may signal, in his words, a "post-normal evaluation" (Schwandt, 2019). If evaluation is to adapt to a post-normal world, the practice and the profession will have to change too. Evaluators will have to be capable of assimilating the assumptions mentioned in the previous paragraph to address structural inequalities. At the same time, evaluation must encourage knowledge co-creation and move beyond methodology to show value sustainability. And it will have to accomplish all this in the "post-truth" and anti-science world described by Patton's contribution. Stronger approaches to knowledge co-construction and active research will pose new challenges for monitoring and evaluation.

Exceptional circumstances challenged expectations and assumptions about what evaluators look at and how they work (Barrados & Lonsdale, 2020). When the pandemic began, numerous reports and studies were quickly produced dealing with a wide range of evaluation topics to respond to a "seismic shift" in ways of working (Buchanan-Smith, 2021, p. 7). Many of them provided operational guidance that reiterated basic evaluation principles and adapted practices to an emergency context (Buchanan-Smith, 2021). The Australian Evaluation Society, for example, offered the following adjustments to practices, including through:

- Reassessing objectives: Updating evaluation objectives to ensure they remain useful.
- Shifting phasing: Changing delivery timeframes and milestones.
- Adapting design: Shifting design, methods, and data collection to achieve the evaluation's objectives.

- Appropriately engaging stakeholders: Considering how COVID-19 is affecting key stakeholders and adapting engagement methods appropriately.
- Contextualizing findings: Interpreting data and forming findings based on contextualized information across different phases of the crisis (e.g., the response and recovery phases) (Australian Evaluation Society Relationships Committee, 2020).

But beyond the technological advances discussed in the previous section, evaluators have had to engage in a process of "sense-making" to ensure that meaningful knowledge is captured on the ground, with substantive and representative information collected from stakeholders. This suggests in turn that the pandemic has served as a window of opportunity to enhance context-sensitive evaluation and country-led evaluation through local consultants, even after the pandemic. In this spirit, various entities have encouraged collaboration and joint evaluations to leverage learnings and improve efficiency.

One of them was the Canadian public service, which looked at the process of "sense-making" through institutional processes and operational changes to enhance collaboration. Authors Barrados, Montague, and Blain in their chapter in this volume, "COVID Crisis – Time to Recalibrate Evaluation," examine a case study in the Canadian government where evaluators teamed up with in-house auditors to create new and effective synergies. Forward-looking approaches like these integrate substantive and higher-level multidisciplinary perspectives and require both creativity and inclusivity.

Evaluation, turbulence, and substantive transformation

A few months into the pandemic, it quickly became apparent that more was going on for evaluators than operational changes resulting from lockdowns, travel bans, or waiting for "normal" to resume. Evaluation was asked to be responsive and nimble to contribute to planning, programme design, and implementation during an evolving series of shocks. Evaluators had to provide data and evidence to inform decision-making quickly. And they were also forced to shift their gaze to higher-order thinking that rises above individual studies and the internal logic of particular interventions to engage with new realities.

In turbulent times, which of course include more pointed and critical junctures like emergencies, evaluation, and its approach to knowledge production have the potential to offer approaches distinct from other disciplines and fields of study (Furubo et al., 2013). New problems required new and more flexible courses of action with shorter horizons for decision-making, while choices about which disciplines, and fields

of knowledge matter became critical. None of these observations are new, but together they appear to have assumed greater significance during the crisis.

Jan-Eric Furubo's chapter, "What Does the Pandemic Mean for Evaluation?" opens the volume with an exploration of why evaluators did not shift their gaze on an urgent basis to the kind of substantial transformation advocated by Patton (Patton, 2021). People and institutions often have sluggish responses or even inertia in the face of crisis. Furubo argues that difficulties in truly perceiving the scale and implications of a crisis can lead to reliance on "what worked" in the past to justify decisions. Furubo discusses the literature on critical junctures and punctuated points of equilibrium that render people unable to acknowledge the scope and scale of change. He offers two important insights in this regard. First, policymakers' original perception of a situation has a direct effect on *which* knowledge is relevant to policy responses. Second, Furubo observes that there is limited capacity to search for new forms of knowledge, especially when those forms of knowledge contradict received wisdom or established policy pathways.

Instead, decision-makers continued to rely on policies and institutions that had continued beyond what Furubo refers to as their "best-before date." Furubo notes that fear of change, path dependency, and the financial costs associated with change can all push decision-makers to adopt similar behaviours. Evaluation systems are supposed to build knowledge iteratively, leading to improvements, but they often repeat and reflect the same explicit or implicit assumptions that underpinned previous or original decisions about the intervention (Leeuw & Furubo, 2008). Furubo's observations are supported by studies of these phenomena, including a tendency to institutional isomorphism, which can be observed in human systems (Ashworth et al., 2009; DiMaggio & Powell, 1983).

Moving away from the past to address fundamental transformations requires an acknowledgement of the changes themselves and substantive knowledge of the challenges before us. The idea of transformation and what it means for evaluation is described by the Independent Evaluation Group of the World Bank, as "an intervention or series of interventions that helps to achieve deep, systemic and sustainable change with large-scale impact in an area of major development challenge" (World Bank, 2016, p. 1).

Barrados, Montague, and Blain illustrate in their chapter how quickly "business as usual" unravelled at the outset of the pandemic. Their chapter also reveals how the institutional inertia described by Furubo can be overcome and allow for quick pivots and genuine transformation. The authors describe a case study involving the Public Health Agency of Canada (PHAC), the events of which unfolded after Canada's federal parliament asked for an assessment of the federal government's response to the pandemic.

The Auditor General of Canada responded in a report that indicated that the government had essentially been unprepared for the pandemic (Office of the Auditor General, 2021). There was widespread media coverage of the report, resulting in criticism directed at the PHAC, placing enormous pressure on evaluators in PHAC to change their approach and establish their value proposition. In rapid time, a pivot was achieved through three main strategies. First was a series of operational innovations to support management in dealing with difficult problems, including by providing real-time support to respond to the COVID-19 crisis. These supports included the mobilization of skills, capacity, and leadership through dedicated teams that assumed responsibilities for incident management in the various areas of work that were directed to addressing COVID-19. These efforts included direct support to the Chief Public Health Officer. Second, PHAC took more advantage of the potential synergies between evaluation and audit functions to provide deeper insights through exchange and collaborative work. Third, the team proactively defined what success looks like in the near, medium, and longer terms. By operationally defining and assessing key performance measures, early warning systems of programme problems were identified through leading and lagging indicators.

This case study illustrates an interesting transformation that took place in almost real time during the pandemic to support decision-makers during times of turbulence. Simultaneously, there were important shifts of another kind, as demands for knowledge production and co-production were also transformed. The alternative to co-production and engaging with multidisciplinary approaches may well result in the declining relevance of evaluation. According to Naidoo in his chapter, there was a rapid incursion of other knowledge actors during the pandemic who moved in quickly to assert their influence. In some cases, these other knowledge actors showed greater responsiveness and agility in the evaluation space, ultimately challenging evaluation's role.

A case in point is the chapter, "The Role of Evaluative Information in Parliamentary Oversight of the Australian Government's Responses to the Pandemic." Peter Wilkins shows how evaluation practice appeared to be absent at the government level in terms of its capacity to play a real role, in real time, in responding to the pandemic. The Australian parliamentary committee responsible for overseeing the government's responses to the pandemic encountered difficulties in accessing information about the pandemic, the measures taken to respond to it, and its impact. In this context, there were important opportunities for the evaluation community to step up and fill some of the gaps. And while it is true that evaluative *information* played a role in the deliberations of the Australian parliamentary committee, Wilkins reaches the startling conclusion that evaluation *per se* did not play an obvious role. In fact, evaluation information was not recognizable in terms of formal

evaluative methods or assessment of the quality of the information. The challenges to formal evaluation as a practice need to be recognized and faced, and Wilkins posits that those challenges may explain why the evaluation community was not more visible. Having said that, the committee's inquiry did receive evaluative information through other sources, including from the Auditor General which provided evaluative information to Parliament which was available for the committee's use.

The pandemic also had the effect of deprioritizing evaluation in many contexts as resources were desperately and immediately needed elsewhere, including for public health and fiscal stabilization. Decision-makers were faced with new problems and had to identify courses of action rapidly, based on reliable information about which disciplines, or fields of knowledge, were relevant in given situations, a point made by Furubo. As well, the practical limitations on evaluation practice (e.g., decreased mobility and limited onsite access) had a significant impact on data collection and enhanced the importance of alternative forms and sources of information. Finally, and more fundamentally, there was a need to revisit evaluation plans and methodologies, and to engage with new policy priorities.

The pandemic also offered insights into the oversight and accountability architecture of countries and institutions. Naidoo discusses the disruption caused by the COVID-19 crisis and argues that it accelerated changes to evaluation's ecosystem from international and regional organizations to nation states and donors, and from evaluation professionals to communities. As well, there appear to have been shifts from agency-specific and vertical modes of working towards more collaborative and horizontal modalities. As the bandwidth for evaluating discrete programmes narrowed, the relevance of specialized approaches may have decreased as well, placing a renewed emphasis on substantive judgement, accountability, and transparency. As Naidoo points out in his chapter, the traditional focus on discrete interventions may have little to no value in an all-of-government or all-of-society approach. Evaluators were forced to become more proactive and deliver results more quickly, raising the possibility of more *ex-ante* work, modelling and, potentially, more reliance on big data.

The important role of audit institutions is also made in the chapter in this book by Jeremy Lonsdale and Maria Barrados. The authors show how state audit institutions in the UK and Canada, in contrast to many traditional evaluation offices, were able to pursue timeliness and flexibility in knowledge production to be useful and valued *during* the COVID-19 crisis rather than providing information after the fact. Building on their previous work in this area, Lonsdale and Barrados acknowledge the differing approaches and perspectives of audit and evaluation. But they also underscore common features and significant crossovers in practice that offer cross-disciplinary learning and exchange, citing their earlier work on this topic (Barrados & Lonsdale, 2020). In particular, the core work of

state audit institutions – performance audits, investigations and "lessons learned" outputs – provided information and knowledge to decision-makers; these institutions were able to do so at pace in environments that were characterized by crisis, rather than recording and rendering an account to others after the fact. The first function has become the highest priority for governments in both Canada and the UK, especially for those working on the "front lines" of the pandemic. This insight serves an important purpose if societies are to learn from the experience of the pandemic.

Systems thinking and complexity

Conventional results-based management, with its linear theories of change and pre-determined impacts, cannot contribute meaningfully to shared responsibility for results. Systems thinking and complexity science present an alternative to these approaches, and instead encourage cross-institutional, cross-system, and cross-country evaluations. They also illustrate the need to shift from individual studies to knowledge streams. Moving from studies to streams has been a prominent theme in evaluation studies and is relevant not only to knowledge production but also to the future deployment of that knowledge (Rist & Stame, 2011). Simply put, evaluating in complex circumstances differs in important ways from linear and more static models of interventions and evaluation (Bamberger et al., 2015; Patton, 2011, 2020).

And yet, the barriers to more integrated flows of knowledge appear to be just as high as ever. Institutional arrangements, where every unit or entity has its own salaries, internal structures, and particular evaluation agreements, stand in the way of "working as one," to borrow UN terminology. Universities, for example, have indirect cost rates that in the aggregate can mean that more money is spent on those rates than on the research itself, making co-construction of knowledge difficult. Few evaluators or commissioners have really addressed the practical question of how disparate researchers and institutional units can work together effectively beyond a particular evaluation study or programme. There is a danger that the current environment, with its decreased resources and the added pressure of assessing emergency responses, will result in us falling back on individual studies, to the detriment of the future of evaluation.

Pawson's chapter "Do Lockdowns Work? Evidence from the UK," provides a window into complexity through a different angle, namely the attempt to assess the UK's virus management programme. His contribution offers insights into the challenges confronting evaluators who are called on to examine complexity and systems thinking across the social sciences and their impacts. In response to the pandemic, Pawson points to literally hundreds of interventions that were introduced in what he calls an unprecedented exercise in social control in a complex systems

framework. His work draws on research in psychology, public health, public policy, economics, and law enforcement, among others (although interestingly, very few specialists in evaluation). Pawson argues that the sheer complexity of these interventions generated scores of emergent and unanticipated effects, requiring revision after revision to lockdown policy. Based on his research, he examines seven classic system dysfunctions that are based on primary evidence to show how the underlying policy assumptions became destabilized.

Pawson's chapter raises similar questions to those raised by other contributors in this volume such as Naidoo, Furubo, and Wilkins, namely, that evaluators were ill-prepared to respond to these complex systems and the real-time demands of the circumstances. The result was that knowledge was sometimes chosen from other sources. The primary research evidence moved directly, if not perfectly, from other knowledge providers to decision-makers, bypassing evaluators more or less completely. If the old ways of working no longer work, attention must be paid to new ways of working. As argued by Lonsdale (2020), academic research and public sector practice have observed the importance of recognizing that one discipline will rarely be enough to address complex public policy problems. Individual lines of inquiry can be enhanced by sharing insights from other disciplines and sources. For example, internal audit and evaluation functions are required in Canadian government departments: even before the pandemic, the potential for sharing and collaboration between performance audits practiced by external audit offices, evaluation, and evaluation and internal audit had been noted (Barrados & Lonsdale, 2020).

Efforts to create greater synergy between audit and evaluation in Canada started in the 1990s, albeit with limited success. Barrados, Montague, and Blain build on that work in their chapter by illustrating how the pandemic presented opportunities to deepen the exchange of practices and foster greater collaboration during crisis in the Canadian federal public service. The case study showed how audit and evaluation were able to work together more effectively and gain from the shared expertise. While the Canadian example showed the importance of harnessing talent and leadership to manage both these functions together, it also showed a positive example of how co-development of knowledge led to a deeper understanding of programmes and offered opportunities for methodological innovation and increased efficiencies.

Evaluators are, or should be, in a unique position to develop and co-create knowledge and to engage in active research that goes beyond the isolated "expert evaluator" working on a single project or study, even if there are still considerable barriers to co-creating knowledge. This approach, in turn, requires us to synthesize and divert streams of knowledge into coherent and usable content-based information that can transcend particular methodologies and technical approaches, while retaining

the evaluation-specific objectives of supporting good governance and accountability. One of the most significant areas for this exercise is the SDGs, which for many years have framed development efforts and national evaluation planning.

Evaluation, the pandemic, and the SDGs

The SDGs had served as common indicators to measure progress, but the magnitude of the crisis resulted in changes that fundamentally affected established systems, including evaluation. Established practices around evaluation had been gaining momentum prior to the pandemic and were increasingly linked to the SDGs, which required progress-reporting among other accountability tools. Key evaluation events demonstrated the positive role that evaluation played in SDG attainment.

In 2020, established practices of reporting national progress against set plans were paused, and the global meta-framework of the SDGs was destabilized. Some of the progress and development gains that had been achieved prior to 2020 were either stalled or in retreat, with system-wide interconnections frayed or broken. The key areas of concern are well-known: the environmental-human development nexus; the capacity for systems thinking and complexity; shared responsibility for results; and a focus on sustainability and human-rights based approaches that integrate core and universal human rights norms into decision-making. The SDG setbacks during the pandemic have been especially evident in the areas of poverty alleviation and gender equality. The resumption of robust economic activity, particularly industrial activity and transportation, has meant that any moderate environmental gains that may have been made during the pandemic have been quickly reversed.

Despite these setbacks, the SDGs, if not the timing for specific targets, remain as relevant today as ever. The fact that many, if not all the SDGs, are now unlikely to be achieved within the projected time frames begs the question of what is next for human development and how it will be measured. At the same time, the crisis offered the opportunity to better understand the capacities of new and existing responses as well as coordination among systems. If the pandemic is indeed an environmental crisis at its core, evaluation must be capable of providing evidence from the real world about the close interlinkages between ecosystem health and human health. Some organizations are viewing the crisis as an opportunity to generate innovation and momentum for the SDGs and the 2030 Agenda as implementation plans are revisited (IEO/UNDP & OECD/DAC, 2020).

At a minimum, the targets and timing of the SDGs will likely have to be rethought in order to reflect more realistically the impacts of the pandemic and translate them into the "post-normal" evaluation world. Robert Lahey and Dorothy Lucks in their chapter, "The Impact of the

COVID-19 Pandemic on the Effective Use of Evaluation in Supporting the Sustainable Development Goals" discuss how the COVID-19 pandemic overtook the previous focus on targets and indicators. It placed an imperative on the evaluation sector to consider how evaluation theory and practice should respond to the effects of the pandemic to ensure ongoing relevance, effectiveness, and efficiency in supporting SDG achievement. Lahey and Lucks ask what COVID-19 meant for the use of evaluation in supporting the implementation, management, and reporting of the SDGs and how evaluation practitioners adapted to ensure that the SDGs stayed relevant and responded to country-level needs, given the COVID-19 context. Using their SDG Monitoring, Evaluation and Learning Framework, the authors make inferences as to where, when, and how the COVID-19 pandemic had impacts on the role and use of evaluation in supporting the SDGs at the country level.

The global pandemic may have raised both awareness and a sense of urgency among leaders of the need to deal with threats that are global in nature, all of which are reflected in the SDGs. As high-level objectives, Patton emphasizes in his chapter the importance of the interdependence of equity and sustainability as intersecting and mutually reinforcing criteria, and the SDG principle of "leaving no one behind" means that there are also needs to be a substantial transformation in the approach of evaluation to human rights. Amnesty International has described this transformation as supporting "a human rights-centred transition to a green economy" as a priority (Dubb, 2020).

Human rights-based approaches

A human-rights centred transition necessarily requires a human rights-based approach to evaluation. Human rights matter to evaluation practice. This is not just because of the imperative of lawfulness in evaluation practice and ethics, but because human rights form part of the global order, starting with the UN Charter.

The systemic weaknesses that were exposed during the pandemic may have been present before it, but their impacts exposed or exacerbated levels of vulnerability, threat, and scarcity. These outcomes have profound implications for evaluation practices because evaluators will be required to assess the impacts of measures taken to address the pandemic in relation to the gaping social fissures that have opened and to track them against the costs to human lives and human rights.

In her chapter, "The Unbearable Lightness of Rights: Evaluation and COVID-19 Responses," Eliadis reflects on these developments and discusses the disproportionate and often lethal impacts on populations that were already under strain. These groups include racialized people and especially Black and Indigenous populations, as well as migrants and refugees,

prisoners, people with disabilities, older people, and those who have had no choice but to continue working in jobs that expose them and their families to infection. Ensuring the alignment of emergency responses to human rights standards means that evaluators must be familiar with substantive human rights standards.

And yet, although evaluators should have been first in line to notice the magnitude of the human rights issues that the pandemic ushered in, Eliadis argues that human rights-based approaches have rarely been centred within evaluation practice, even in the post-crisis or post-normal era. She underscores the weak role that human rights play in practice in most of evaluation's ethical frameworks, building on earlier work in this evaluation series (van den Berg et al., 2022). She argues for a substantive strengthening of both human rights and human rights-based approaches as central load-bearing beams in evaluation practice. Finally, the chapter examines what evaluators need to know about the particular rules that apply to times of emergency in order to ensure compliance with the principle of legality.

Conclusion

The environmental-human development nexus relies on medical advice and on "the science," as well as on a wide range of public health, social science, and legal disciplines. All require a contextual understanding of the zoonotic origins of the virus and of the wide impacts on environmental and human systems. This does not mean that evaluators must become scientists, economists, or lawyers. However, evaluators who ignore interdisciplinary approaches risk perpetuating or repeating the conditions that gave rise to many of the policy failures that took place during the crisis. It is incumbent on evaluators to inquire into what we can learn about the role of evaluation in such situations, since it is unlikely that this pandemic will be the last such event. Certainly, paying attention to lessons learned may improve our resilience and preparedness for future systemic shocks (Independent Evaluation Group, 2017).

The contributions in this volume invite us to draw lessons from the past and to think prospectively about how recovery will be managed, and about the transformation of evaluation that will be needed to support it. In his Afterword, Ray C. Rist points to five observations that evaluators must consider that go far beyond technical and operational changes to evaluation practice: first, "big government" had the major role to play in the managing the crisis, a role that markets could never have played; second, gross inequities were ignored until far too late and must be addressed if the social contract is to be renewed with the most marginalized in our society; third, and linked to the previous point, unequal and often inaccessible health systems were both a cause and a consequence of

the pandemic's human impacts; fourth, misinformation and conspiracies played an outsized role in how the pandemic was perceived and managed, and fifth, there is no going back to "normal." In the face of possible new waves of infection, new or reinforced emergency measures or entirely new pandemics, we ignore what we have experienced and learned at our peril.

Notes

1 Many thanks to Mariel Arumburu who conducted the research on the meta-analysis of evaluation literature published in the first year of the pandemic, and to Ms. Arumburu and Peter Wilkins for comments and input on an earlier draft. The editors are grateful to Elizabeth Fraser for her thorough research assistance in the project.
2 African Development Bank, Independent Development Evaluation; Asian Development Bank, Independent Evaluation; Inter-American Development Bank, Evaluation and Oversight and the World Bank, Independent Evaluation Group.
3 UN Development Programme (UNDP), Independent Evaluation Office; UN Evaluation Group (UNEG), UN Women, IFAD Independent Office of Evaluation.
4 African Evaluation Association (AfrEA); Asia-Pacific Evaluation Association; Caribbean Evaluators International (CaribEval); Community of Evaluators (CoE) in South Asia; Eurasian Alliance of National Evaluation Associations (EvalEurasia); European Evaluation Society; Evaluators Network of the Middle East and North Africa (EvalMENA); Red de Seguimiento, Evaluacion y Sistematización en America Latina y el Caribe (ReLAC 2.0).
5 Australian Evaluation Society; American Evaluation Association; Canadian Evaluation Society; Ghana Monitoring and Evaluation Forum; Lebanese Evaluation Association (LebEval); Academia Nacional de Evaluadores de Mexico (ACEVAL); South African Monitoring and Evaluation Association (SAMEA); Réseau Tunisien d'Evaluation (RTE).
6 Asian Development Bank, Independent Evaluation; Inter-American Development Bank, Evaluation and Oversight and the World Bank, Independent Evaluation Group; UN Development Programme (UNDP), Independent Evaluation Office; UN Evaluation Group (UNEG), UN Women, IFAD Independent Office of Evaluation, the European Evaluation Society, the Australian Evaluation Society, the American Evaluation Society, and OECD (DaC EvalNet)/UNDP. We also undertook a scan of COVID-19 related publications in major evaluation journals, evaluation society journals, and major publishers, but given the compressed timeframe and the delays inherent in peer review, those publications were not reviewed for the purposes of this discussion.
7 Although at the same time, the Asian Development Bank stressed the independence of evaluation: to remain useful, effective, and relevant in times of crisis, evaluation had to continue playing its role in both its functions of accountability and learning, without compromising its independence (Salze-Lozac'h, n.d.).

References

Arumburu, M. (2021). "Meta-analysis/web site survey of evaluation key trends and learnings during COVID-19" (Unpublished). On file with authors.

Ashworth, R., Boyne, G., & Delbridge, R. (2009). Escape from the iron cage? Organizational change and isomorphic pressures in the public sector. *Journal of Public Administration Research and Theory*, *19*(1), 165–187. https://doi.org/10.1093/jopart/mum038

Australian Evaluation Society Relationships Committee. (2020, 29 June). Statement on evaluation during the pandemic. *AES Blog*. https://aes.asn.au/aes-blog/statement-on-evaluation-during-the-pandemic

Bamberger, M., Vaessen, J., & Raimondo, E. (Eds.). (2015). *Dealing with complexity in development evaluation: A practical approach*. Sage.

Barrados, M., & Lonsdale, J. (Eds.). (2020). *Crossover of audit and evaluation practices: Challenges and opportunities*. Routledge.

Buchanan-Smith, M. (2021). *Learning as we go: How the Pandemic is changing evaluation strategy and planning*. ALNAP. https://www.alnap.org/system/files/content/resource/files/main/ALNAP%20paper%20-%20Learning%20as%20we%20go.pdf

Center on Gender Equity and Health, UCSD, Data2X, Bill & Melinda Gates Foundation, Global Center for Gender Equality at Stanford University, Global Health 50/50, International Labour Organization, PARIS21, UN Women, & World Bank. (2021). *Strengthening gender measures and data in the COVID-19 era: An urgent need for change*. UN Women. https://www.unwomen.org/sites/default/files/Headquarters/Attachments/Sections/Library/Publications/2021/Strengthening-gender-measures-and-data-in-the-COVID-19-era-en.pdf

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review, 48*(2), 147–60. https://www.jstor.org/stable/2095101

Dubb, S. (2020, January 16). *Amnesty and Planned Parenthood identify their top policy priorities*. Nonprofit Quarterly. https://nonprofitquarterly.org/amnesty-and-planned-parenthood-identify-their-top-policy-priorities/

Forss, K., Lindkvist, I., & McGillivray, M. (2021). *Long-term perspectives in evaluation increasing relevance and utility*. Routledge.

Furubo, J.-E., & Stame, N. (Eds.). (2019). *The evaluation enterprise: A critical view (comparative evaluation studies)*. Routledge.

Furubo, J.-E., Rist, R., & Speer, S. (2013). *Evaluation and turbulent times: Reflections on a discipline in disarray*. Transaction Publishers.

Global Evaluation Coalition. (n.d.). *About*. https://www.covid19-evaluation-coalition.org/about/

Hassnain, H. (2020). Mitigating the risks of remote data collection for evaluations. *EarthEval*. https://www.eartheval.org/blog/mitigating-risks-remote-data-collection-evaluations

Independent Evaluation Group. (2017). *Crisis response and resilience to systemic shocks: Lessons from IEG evaluations*. World Bank. https://ieg.worldbankgroup.org/sites/default/files/Data/reports/building-resilience.pdf

Independent Evaluation Group. (2020). *Results and performance of the World Bank group 2020: An independent evaluation*. World Bank https://openknowledge.worldbank.org/handle/10986/34844

Independent Evaluation Office/United Nations Development Programme & Organisation for Economic Co-operation and Development/Development Assistance Committee. (2020, April). *Good practices during COVID-19*. United Nations Development Programme. http://web.undp.org/evaluation/guideline/documents/covid19/IEOOECD_DAC_Joint-Guidance_COVID19.pdf

Independent Office of Evaluation. (2021). *The experience of the Independent Office of Evaluation of IFAD in conducting evaluations during Covid-19*. International Fund for Agricultural Development. https://www.ifad.org/en/web/ioe/w/the-experience-of-the-independent-office-of-evaluation-of-ifad-in-conducting-evaluations-during-covid-19-learning-note

Leeuw, F. L., & Furubo, J.-E. (2008). Evaluation systems what are they and why study them? *Evaluation*, *14*(2), 157–169. https://doi.org/10.1177%2F1356389007087537

Lonsdale, J. (2020). Conclusion. In M. Barrados & J. Lonsdale (Eds.), *Crossover of audit and evaluation practices: Challenges and opportunities* (pp. 208–219). Routledge.

Office of Evaluation and Oversight. (n.d.). COVID-19: What can we learn from past responses to public health crises? *Inter-American Development Bank*. https://www.iadb.org/en/ove/covid-19.

Office of Internal Oversight Services. (2020, June 12). *Synthesis of guidelines for UN evaluation under COVID-19*. United Nations Evaluation Group. http://www.unevaluation.org/document/detail/2863.

Office of the Auditor General. (2021). *Report 8 – Pandemic preparedness, surveillance, and border control measures*. https://www.oag-bvg.gc.ca/internet/English/parl_oag_202103_03_e_43785.html

Office of the Privacy Commissioner of Canada. (2022). *2020–21 Departmental Results Report*. https://www.priv.gc.ca/en/about-the-opc/opc-operational-reports/reporting-of-opc-spending/drr_index/2020-2021/operating-context_2020-21_drr/

Patton, M. Q. (2011). *Developmental evaluation: Applying systems thinking and complexity concepts to enhance use*. Guilford Press.

Patton, M. Q. (2020). *Blue marble evaluation: Premises and principles*. Guilford Press.

Patton, M. Q. (2021). Evaluation criteria for evaluating transformation: Implications for the coronavirus pandemic and the global climate emergency. *American Journal of Evaluation*, *42*(1), 53–89. https://doi.org/10.1177/1098214020933689

Petersson, G. J., & Breul, J. D. (Eds.). (2017). *Cybersociety, big data and evaluation*. Routledge.

Raimondo, E., Vaessen, J., & Branco, M. (2020). Adapting evaluation designs in times of COVID-19 (coronavirus): Four questions to guide decisions. *Asian Development Blog*. https://ieg.worldbankgroup.org/blog/adapting-evaluation-designs-times-covid-19-coronavirus-four-questions-guide-decisions

Rist, R. C., & Stame, N. (2011). *From studies to streams: Managing evaluative systems*. Routledge.

Salze-Lozac'h, V. (n.d.). COVID-19 projects and programs need careful oversight. *Asian Development Bank*. https://blogs.adb.org/blog/covid-19-projects-and-programs-need-careful-oversight

Schwandt, T. A. (2019). Post normal evaluation? *European Evaluation Journal*, *25*(3), 317–329. https://doi.org/10.1177%2F1356389019855501

United Nations. (2020. April). *COVID-19 and human rights: We are all in this together.* United Nations Sustainable Development Group. https://unsdg.un.org/sites/default/files/2020-04/COVID-19-and-Human-Rights.pdf

United Nations Population Fund. (2020). *Adapting evaluations to the COVID-19 pandemic. Guiding principles*. https://www.unfpa.org/sites/default/files/admin-resource/FINAL_Covid19_and_Eval.pdf

van den Berg, R. D. (2022). Ethical guidance: Enabling evaluation's role in tackling bad and doing good. In R. D. van den Berg, P. Hawkins, & N. Stame (Eds.), *Ethics for evaluation beyond "doing no harm" to "tackling bad" and "doing good"* (pp. 17–44). Routledge.

van den Berg, R. D., Hawkins, P., & Stame, N. (Eds.). (2022). *Ethics for evaluation beyond "doing no harm" to "tackling bad" and "doing good"*. Routledge.

van den Berg, R. V., Magro, C., & Adrien, M.-H. (Eds.). (2021). *Transformational evaluation for the global crises of our time*. International Development Evaluation Association.

World Bank. (2016). *Supporting transformational change for poverty reduction and shared prosperity: Lessons from World Bank group experience*. https://documents1.worldbank.org/curated/en/350031467991920112/pdf/Supporting-transformational-change-for-poverty-reduction-and-shared-prosperity-lessons-from-World-Bank-Group-experience-an-IEG-category-two-learning-product.pdf

1 What Does the Pandemic Mean for Evaluation?

Jan-Eric Furubo

Since H.G. Wells wrote *The Time Machine* in 1895, books and movies have described how time travelers can become trapped in time. Outside of fiction, societies, institutions, and policies can also be trapped in the present and the past. In periods of relative stability, it is easier to continue a current course of action, interpreting new information within the paradigm of an existing understanding of the world than it is to change more fundamentally. Things change only gradually and incrementally. However, at some point, phases of stability will end, our understanding of reality will be shaken, and a re-orientation of institutions and policies will become inevitable. The literature on societal change, or on historical developments more broadly, uses terms like critical junctures, punctuated equilibrium, and breaking or turning points. Such theories help us to understand why policies and institutions often are not questioned, even if it seems obvious that they have passed their best before date or even their expiry date. It is expensive to change. Change creates uncertainty and it alters dynamics among players. In his discussion of developmental trajectories, Paul Pierson (2004) notes that "… the *relative* benefits of the current activity compared with once-possible options increase over time. To put it in a different way, the costs of switching to some previously plausible alternative rise" (p. 21, italics in original). When an accumulation of gradual changes reaches a tipping point, the need to fundamentally change the present course and the institutional setting of societies and governing structures becomes evident.[1]

The occurrence of such breaking points can also be the result of an overwhelming event or something which we describe as a "crisis." At the time of writing, the global community faced and still faces a crisis of dramatic proportions. We cannot know for certain, but have reasons to believe, that the global, urgent confrontation with a sudden emerging pandemic will have profound consequences. This is something to which this chapter will return, but at this point, it is reasonable to assume that much debate and research will continue to focus on the institutions which in some way had assumed a range of responsibilities in decisions and implementation of actions related to the pandemic.

DOI: 10.4324/9781003376316-2

It can also be assumed that some of the consequences of the pandemic will reach far beyond the policies, institutions, social practices, and professions which, in a more immediate way, have been involved in handling the crisis. This chapter is focused on our own profession; the focus of the discussion will be on how the pandemic will and should impact the practice of evaluation and the understanding of its role in society. The chapter is therefore an effort to discuss our own practice through the lenses of what many of us probably regard as one of the most overwhelming peace time global experiences in generations. The importance of such a discussion has a normative foundation. Practices and professions which claim that their mission is to contribute to a better society must also reflect internally on how such collective experiences will impact the fulfillment of their mission and its role in society.

This ambition makes it clear what this chapter is not about. First, it is not an effort to give preliminary answers, for example, through summarizing early evaluations or about reviewing some aspects of the response to the pandemic. Nor is it about what factors, including social, legal, and institutional conditions and arrangements, might explain differences in how certain specific interventions seem to result in different outcomes or which factors explain the choice of strategies in different countries. Studies focusing on such questions are certainly extremely important and are addressed by several other chapters in this book. We can be certain that in the coming years, we will see many papers and books dealing with these and similar questions.

Second, this chapter will not discuss the fundamental normative questions related to the problems and injustices that the pandemic has revealed. In the Swedish case, for example, it is obvious that one important consequence of what happened in 2020 and 2021 is that the pandemic increased our awareness of shortcomings in how we take care of the oldest in our societies (Ministry of Social Affairs, 2020). The same can be said about the differences in how social and ethnic groups have been affected by the pandemic.

Within these limitations, the chapter is divided in two parts. In the first part, "Interventions, assumptions, and knowledge," three questions are addressed:

1. How can we identify the relevant and often crucial knowledge in policymaking and decisions about interventions?
2. How is the knowledge of evaluation producers different from what is delivered by other knowledge producers?
3. In what way is the role of evaluation and other forms of knowledge different in times of crises than in stable "normal" periods?

The second part of this chapter, "The interface between decision-makers and knowledge producers," will focus on what can be described as

the interface between the decision side and the knowledge side in crises. Using experiences related to the pandemic, I will identify two hypotheses which can, at least partially, explain the underutilization of some types of knowledge during the pandemic:

- The policymakers' *original perception* of a crisis impacts which knowledge will be seen as relevant in designing the response to crises.
- Existing knowledge structures can lead to *asymmetric relations* and the underutilization of relevant knowledge.

Interventions, assumptions, and knowledge

If we want to identify what knowledge is relevant in decisions about and construction of governmental interventions, we can start with a simple statement: governmental policies and programs, or more broadly interventions, exist because some developments are more desirable than others.[2] An intervention can aim to change an existing situation which is seen as a problem or to prevent a development which is seen as undesirable. The decision about the intervention and its construction is based on *assumptions* about which factors determine the future development of, for example, economic growth, the number of suicides, the distribution of resources among the population, greenhouse gas emissions, and so on. This general description of governmental interventions will be the same, irrespective of whether the intervention is one that has been analyzed and prepared over several years or whether it is about the need to react to an urgent problem which emerged only a few days ago.

This does not mean that interventions are solely based on assumptions about how a certain development can be impacted. Values are important too. Even if everyone agrees on the description of a certain situation, it is not certain that everyone will agree on whether the situation is a problem. People can certainly have different opinions about increased wage differences or if it is a problem that significantly more women than men are admitted to prestigious educational establishments. And even if there is an agreement that something is a problem, it is not certain that everyone also agrees that the government should do something about it. The question about when a government should intervene is an important and value-based dividing line in a political debate. In the European Union, as well as in many federal states, there are value-based differences in opinion about which level of government should act in different situations.

Most certainly, even if everyone agrees that something is a problem, and that the government should act, the decision about *how* the government should act is also impacted by values. For example, research in several disciplines shows that physical mistreatment is not a deterrent. However, even if existing research did point in the opposite direction,

many of us would argue against physical mistreatment or capital punishment. Less dramatic examples include discussions around limitations on rights, and, to give an example on a question discussed in many nations during the pandemic, the degree to which temporary reductions of important freedoms (e.g., freedom of peaceful assembly) can be acceptable both to international standards and what we see as fundamental values in a democratic society.

Already, this sketched discussion demonstrates a complex and intertwined relationship between values and assumptions about how a certain development can be impacted. However, as soon as it has been agreed, at least by a majority, that a government should intervene, the assumptions will be important in constructing the intervention. These can be explicit or tacit and they can, of course, be right or wrong. However, governments aspiring to change behavior in some way may decide to use an educational program, and in doing so, are basing interventions on other assumptions about social and psychological causal mechanisms, as opposed to a government which instead prohibits certain forms of behavior.

This discussion about knowledge and governmental interventions, or more generally expressed purposive social actions (Merton, 1936, p. 894), leads to a territory that is well known both in the literature about government policies and in the evaluation literature. Such actions or interventions are based on assumptions about how a chain of events in a causal process concatenate a certain action or institutional arrangement with a certain outcome. This does not mean that these causal assumptions are necessarily supported by earlier research. However, even when they are not, and they may perhaps not even be explicitly expressed, we can talk about assumed causal relations, hypotheses, or theories about such relations. Pawson and Tilley have therefore described programs as "theories incarnate," which begin:

> In the heads of policy architects, pass into the hands of practitioners and, sometimes, into the hearts and minds of programme subjects. These conjectures originate with an understanding of what gives rise to inappropriate behaviour, or to discriminatory events, or to inequalities of social condition and then move to speculate on how changes may be made to these patterns.
>
> (Pawson & Tilley, 2004, p. 3)

In the evaluation literature, this causal chain has often been discussed as the *program theory*, but also, for example, as the *logic model* (Rogers, 2008, p. 30) or *intervention theory* (Vedung, 2009). This chapter refers mainly to the latter term. However, intervention theory can cover different parts of the causal chain between the first, perhaps high-level, decisions about an intervention, to the final results in terms of better health, improved environment, or less poverty. In the evaluation literature, Chen has therefore

made the distinction between the *action model* and the *change model*. The change model covers the causal process, which is generated by the intervention; it is, in other words, about what will happen after the intervention is delivered. Chen argues that the change model consists of three main components: (a) an intervention, which refers to a set of program activities that focus on changing the determinants and outcomes; (b) determinants, which refers to levers or mechanisms that mediate between the intervention and its outcomes; and (c) outcomes, which refers to the anticipated effects of the program (Chen, 1990, 2006, p. 76ff.). When Chen talks about the action model, it is about what happens in the phases before the moment of delivery. It is about the planning of which resources should be used and which organizational entities should participate and interact, how the intervention should reach different groups, and so on (Chen, 2006, p. 76). Many questions related to the discussion about the governmental responses to the pandemic can be connected both to the action model and the change model, and the line between them is not very sharp. However, for the purpose of this chapter, it is helpful to focus on the change model, the causal chain which starts with a certain intervention.

Two groups of assumptions

The assumptions about the causal relation on which an intervention is based, what Chen described as the change model, can in turn be categorized in different ways and across different dimensions. In this chapter, I distinguish between two main groups of assumptions and argue that discussions about the response to the pandemic have been about assumptions within these two groups.

The first group are the *substantive assumptions*, which are related to situational and material problems. These assumptions explain how certain behaviors and measures can change a specific situation; in other words, how a problem can be solved, or its effects reduced, by certain acts. If the government tries to get more cyclists to use helmets (irrespective of how the government does this), it most probably will do so because it assumes that an increased use of helmets by cyclists will lead to fewer serious injuries if the cyclist is in an accident. And why should a government spend money on subsidies to house owners to get them to invest in solar panels if it is not advantageous for the environment? Or why would a government try to influence the sunbathing habits of its population if research had not shown that the risk for malignant melanoma was correlated with sun exposure? Such assumptions about behavioral changes or actions taken by individuals, business enterprises, municipalities, and so on are the starting point for most governmental interventions.

These substantive assumptions are often based on theories regarding scientific, biological, virological, or technical mechanisms, but can also

be about social or psychological processes and mechanisms. Many of these assumptions can also be tested with experimental methods, for example, the causal relationship between liming and acidity, or general decay in a residential area and crime (see, e.g., Leeuw & Nelen, 2013, p. 187ff.).

The second group are *governance assumptions*, related to the choice of governing instruments which can be used to impact behavior and the frequency of certain measures. How can the government impact individuals to reduce their sun exposure, get house owners to invest in solar panels, or municipalities to invest in liming lakes or hiring more qualified teachers? Will some investments be made first when they become more economically advantageous? Or will some actions or behavioral changes take place first when some behaviors or actions are mandatory or prohibited? The answers to such questions have to do with assumptions about the mechanisms which are decisive for why people or organizations act in different ways, or to put it differently, how the diffusion of behaviors and actions can be impacted by different tools of government or what we can describe as policy instruments.[3]

When we want to find evidence for or against such governance assumptions, we often look to fields such as psychology, organizational theory, economics, and diffusion theory. To give an example from the discussion about how to respond to the pandemic, research about trust in institutions can impact governance assumptions. In this group of assumptions, we can also include assumptions about how different factors can interact or oppose each other, such as displacement effects. In some situations, government subsidies of certain investments can suppress others in the same field, while government subsidies to municipal advisers about energy conservation measures can suppress the commercial market in the same field.

The discussion about the response to the pandemic can be related to the above discussion. Nearly all national decision-makers and all relevant international institutions have agreed that the pandemic was an urgent problem. Governments at all levels had a responsibility to cope, one way or another, by limiting viral transmission and mitigating its economic and social effects. There seems to have been some consensus among organizations like the World Health Organization (WHO) and the European Centre for Disease Prevention and Control (ECDC), national authorities and governments, about the importance of physical distancing and other substantive *assumption*s. However, one exception to this picture of unity seems to have been whether wearing face masks affected the spread of the virus.

Most of the debate about the political response to the pandemic appears to have been about how different governing tools should be used to impact behaviors and change the frequency of different measures, the *governance assumptions*. The pandemic is therefore one of many cases showing that differences about policy are often more about differences in *how* a

government should impact a certain situation more than *what* behavioral or material changes are desirable. One illustration is the varying standpoints, discussed in the chapter by Pearl Eliadis, when it comes to question about the degree to which *legal restrictions* are unavoidable, whether they violate human rights standards, and what the response of evaluators should be to this phenomenon. In some countries, Sweden being a pronounced example, it has been argued that it is only possible to maintain behavioral changes in the long term if there is a broad and shared understanding of why they are important. The Swedish authorities have further argued that prohibitions and mandatory controls, when they are applied in many different situations, will be challenged if they seem arbitrary to the public. Legal restrictions should therefore be used rarely and instead; improved education and information can lead to increased knowledge among the public and thereby result in more successful voluntary behavioral changes.[4] However, many other governments have taken the opposite position and instead assumed a causal relation between the degree of legally binding restrictions and decreased transmission. For the purposes of this chapter, the important thing is not to ask which of these positions has the strongest scientific support. What is important is instead that the choice of policy instruments was not discussed in an explicit and transparent way, in terms of the assumptions made about how behavior and actions could be impacted by these instruments, and the degree to which the efficiency of different policy instruments can vary due to cultural and other factors, such as the level of trust in a society.

In this context, it is notable that prestigious scientific organizations could base their analyses on very simplified, and not transparent, assumptions about the efficiency of different governing tools. In an early paper, Imperial College took it for *granted* that legally binding restrictions were the most effective approach to reducing viral transmission. In the forecast, they therefore assumed that the transmission of the virus was strongly correlated with the degree of economic and social activities that were formally and legally restricted by the government (Flaxman et al., 2020, p. 6). However, the underlying assumption for this standpoint was not explicit and no references were given to studies supporting it.

As mentioned above, it is possible to find exceptions to the broad consensus about substantive assumptions. The debate over face masks raises both substantive and governance assumptions. During most of 2020, leading officials within the National Public Health Agency in Sweden argued that even if face masks resulted in decreased transmission of the virus (a substantive assumption), it was possible that wearing face masks could also lead to a false feeling of security (a governance assumption). This latter assumption could reduce many individuals' willingness to distance themselves from others and abstain from taking journeys, for example. The result would be a possible negative net effect (Cederblad, 2020;

Erlandsson, 2021; Hedenvind, 2020). However, other experts who took part in the Swedish debate argued instead that it was possible to assume the existence of a positive additional side effect: people who wear face masks and see others doing so will be more careful when it comes to social distancing (Milstead, 2020).

The discussion demonstrates the value of program and intervention theories in identifying policy-relevant assumptions. To a significant extent, evaluation theorists have been the main contributors to the use of a tool which is very important for our understanding of which knowledge is important in decisions regarding different interventions. In addition, the discussion has hopefully demonstrated the potential usefulness of these analytical instruments, in a situation when decision-makers need to make a more immediate response in a dramatic situation, like the coronavirus pandemic. A further question is whether this potential has been exploited.

Evaluation and other forms of knowledge

As other chapters in this book observe, understanding the role of evaluation in times of crisis may also help to better understand why evaluations seem to have been used only to a limited extent. In this respect, we must discuss the role of evaluation in a broader context and its relation to other knowledge producers.

When evaluation as a social practice was developed in the United States (US), it was created to determine the effectiveness of public and social actions programs (Furubo, 2019, p. 7ff.) or, as Weiss phrased it, "as a means of contributing to the *improvement* of the program or the policy" (Weiss, 1998, p. 4, italics in original). Evaluation was defined as something which was undertaken with reference to some intentional action, as von Riecken previously wrote in 1954 (Riecken, 1972, p. 86). Since it emerged as social practice, an important aspect of evaluation has been that it is something which is done purposefully. Another aspect is whether the action being evaluated must necessarily be something which has already taken place (ex-post), or whether it is also something that is still on the drawing board (ex-ante).

This means that evaluation has been delimited from much of the knowledge that is useful, and perhaps even necessary, when designing different actions. A researcher can, to give an example, study how social, biological, and psychological mechanisms affect how children learn to read. This analysis can occur regardless of whether the researcher was initially motivated by the expectation that the results of the research could be useful in the design of various policy initiatives. However, the fact that a study of phenomena takes place (such as how students' performance is affected by teachers' expectations, why young people commit suicide, what affects our holiday habits or the development of trade between countries, etc.) can be

used in the design of policy measures, does not mean that these studies can be called evaluations. If we were to label all studies that provide useful knowledge for the decisions and implementation of government activities as evaluations, then evaluation certainly would be synonymous with almost all forms of knowledge production.

The research object can often be focused on the mechanisms of different interventions. However, the salient feature of evaluation is that the object of evaluation is not these mechanisms *per se,* but the interventions which intend to impact these mechanisms. Therefore, in decisions about how various societal problems could be handled or resolved, other forms of knowledge outside only those which can be obtained from evaluation are important.

Evaluation is quite a different thing: its focus is often on interventions which have been constructed with the goal of impacting these mechanisms. In this sense, evaluation can also contribute to building knowledge about these mechanisms. Weiss has emphasized that evaluation could contribute to the development of *basic knowledge*. She noted that under "appropriate conditions, it [evaluation] can also contribute to the development of basic knowledge […] Finding out how successful these efforts have been, and why, can lead to discoveries about basic concepts of human behavior and social structures" (Weiss, 1972, p. viii). If one central question for evaluation as a social practice is to what degree evaluation *during* the pandemic has contributed to knowledge which was useful in defining responses to it, Weiss' statement therefore raises another question: has evaluation during the pandemic had the capacity to make the discoveries Weiss wrote about, and thereby contribute to knowledge which can even be useful in the *future*?

How crises are different

An earlier book in this series, *Evaluation and Turbulent Times*, discussed more broadly the role of evaluation in times of tumultuous change, but more specifically, also examines what happens when societies find themselves suddenly facing emerging problems which demand urgent actions (Furubo et al., 2013). The present crisis has forcefully demonstrated that the role and the relative importance of both evaluation and other forms of knowledge during crises need to be further, and more critically, discussed.

Crises are related to threats and danger. When we talk about crises, we imply that it is an event which in terms of scale, novelty, and urgency represents something different to that which we normally confront (Boin et al., 2005, p. 3; Leonard & Howitt, 2009, p. 611ff.). A school or a town can have limited adaptive capacity to deal with something like a fire, but we will not describe it as a crisis on a more aggregated social level (except at the specific school involved). Crises must involve an element of novelty. Even if a routine emergency, like a fire, can be very demanding, it is

also predictable and possible to plan for. Leonard and Howitt point out, "Because of unusual scale, a previously unknown cause, or an atypical combination of causes, responders face novel challenges, the facts, and implications of which cannot be completely assimilated in the moment of the crises" (Leonard & Howitt, 2009, p. 614). When we talk about crises, we also imply uncertainty and urgency. It is sometimes unclear which answer is the most adequate, but the response must be delivered more or less immediately.

Decisions in crises are very different from decisions and policymaking in more stable situations. The decision-makers must face extreme *uncertainties* under extreme *time* pressure. The most obvious difference therefore has to do with time. The processes and different phases, which in "normal" policy development can span several years, must be squeezed into days and even shorter periods. Decisions which are formative for future policy development must be based on limited and uncertain knowledge. The time is short (or even non-existent) for acquiring knowledge, which can confirm or falsify assumptions of which measures and behavioral changes are important in the specific situation. This is also true for the time that is available even to identify the assumptions about how the diffusion of these measures and behavioral changes can be impacted – to recall the two groups of assumptions mentioned earlier.

As discussed earlier, evaluation has been characterized by the idea that it was part of an ongoing process of incremental improvement. It was possible to gradually improve programs through evaluation, and leading evaluators expressed a strong belief that incremental improvements in an arsenal of programs would lead to a better society (e.g., House, 1993, p. 138; Suchman, 1967, p. 1; Weiss, 1972, p. 4). In the ongoing process of adjustments and improvements, evaluation could be built in to deliver knowledge. However, it has been pointed out that evaluation systems often reflect the same assumptions as those explicit or implicit assumptions that were behind the original decision about the intervention (Leeuw & Furubo, 2008). These underlying assumptions are therefore often not questioned, and evaluation can sometimes confirm assumptions which should have been disproved if they had not been taken for granted in the evaluation system.

Crises shake the *foundation of different policy interests and structures*. They ask different questions whose answers are not immediately found within the existing policy framework or, as Boswell put it:

> Policymakers are more likely to recognize gaps in research where they become aware of the emergence of new types of problem, such as climate change, the impact of new technologies, threats to public health or security, or the emergence of new forms of criminality or social pathology.
>
> (Boswell, 2009, p. 243)

In the decision about how new problems should be handled, the questions are more open and are much more about how to identify which knowledge is the most relevant given the situation, regardless of whether the knowledge has been gained in the context of earlier interventions.

The important question, then, is not to find out which knowledge has been produced within an existing policy framework. It is rather about a search, in a much more open manner, about the knowledge that exists about *what* we need to know and *who* can answer the questions needed to determine the best course of future actions.

Crises therefore force decision-makers to leave the familiar terrain of earlier policy frameworks. In the discussion about new priorities and actions, knowledge produced within earlier policy frameworks is less relevant. In crises when decision-makers need to test different assumptions about psychological, economic, and other mechanisms, which can impact the behavior and actions of individuals or institutions, they will not ask questions about "which evaluations have been done which can be relevant just now?" They will instead ask questions like "what does the research say about how we can affect the mechanisms that can help solve the problem that must be handled?" In other words, what becomes important is knowledge that may not necessarily be associated with evaluation at all.

The interface between decision-makers and knowledge producers

Even before the WHO's decision to declare the coronavirus outbreak to be a pandemic, many countries were demanding a governmental response at national and other levels. On the international level, the WHO identified the global strategic objectives for reaching the overarching goal of controlling the pandemic by slowing down the transmission and reduction of mortality associated with COVID-19 (WHO, 2020a, 2020b, p. 5). The WHO pointed out that all sectors in society must be mobilized urgently, sporadic cases and clusters must be controlled, and mortality must be reduced by providing appropriate clinical care, protecting vulnerable populations and by developing safe vaccines. All these objectives demand different forms of political action and involvement of political and administrative structures (WHO, 2020b, p. 5).

However, the objective which most directly addresses classical trade-offs and policy problems is the objective to suppress "community transmission through context-appropriate infection prevention and control measures, population level physical distancing measures, and appropriate and proportionate restrictions on non-essential domestic and international travel" (WHO, 2020b, p. 5). The discussion about harder and softer regulations, and the levels of "lockdown," relate to this objective. It became clear quickly that different governments chose varying strategies in relation to

this objective. It was also clear that to a great extent, the sometimes-heated debate about different government policies has been about these strategies, both in the long- and short-term; discussions and decisions have been about which policy instruments or tools of governance will impact the behaviors and economic and social activities in a society in such ways that the transmission of the virus will be suppressed. And the discussion has also been about possible negative and positive side effects of different interventions, which have been related to both other public health issues and, for example, economic considerations.

It is obvious that the knowledge that has proved to be pertinent to the discussions about pandemic response is strongly related to what has been described here as the second group of policy-relevant assumptions. However, the degree to which the body of knowledge and theories relevant for discussing these assumptions has been channeled into different decision-making processes, and whether existing research, can certainly be questioned. This interrogation would have been important in falsifying or confirming crucial assumptions, and in determining adequately use.[5] It appears that one of the biggest challenges in modern times – the task of achieving an impact on population behavior, if only to a limited degree – has been based on evaluations and other forms of knowledge about the possible advantages or disadvantages associated with the instruments chosen by and with the execution of different measures.

The question is therefore whether it is possible to find plausible explanations for such an underutilization of relevant knowledge, which probably includes many evaluations. I will, in the following discussion, point out two hypotheses which can help to explain this phenomenon.

Hypothesis one: The importance of the original perception

The first hypothesis is that the initial perception of the crisis will be decisive for which knowledge will be viewed as the most relevant by decision-makers.

When decision-makers face a new problem, they try to understand "what it is about." This kind of original perception will impact their identification of what knowledge is needed. In more stable situations, when the society and different groups of decision-makers become aware of a new problem, they have ample time to identify and to understand which knowledge is relevant in responding to the problem. In crises, the understanding of the situation must be more immediate. If it is a catastrophic reactor meltdown in a neighboring country, the *preunderstanding* will most probably be that the knowledge needed to handle the situation and mitigate its consequences has to do with questions related to the radiation. Important questions will naturally include how soon the radiation reaches their own country, how long it will stay, how the population can reduce

the risk for exposure, which food should be avoided, and so on. Similarly, when confronted with contagious diseases and we want to reduce the spread of infection, decision-makers will ask for data from knowledge fields such as virology, epidemiology, and public health. The initial perception of a certain crisis defines which knowledge fields and which expert structures are most relevant for the handling of the crisis. This initial perception can therefore also be an obstacle for finding relevant knowledge. This leads to some questions related to the institutionalization on both the demand and supply side of the use of knowledge in policy processes.

Regardless of which policy different governments might adopt, it seems that the knowledge they ask for is dictated by their initial perception of the character of the *present* crisis. This perception and a "quick" path dependency hamper the utilization of knowledge related to governance assumptions. In other words, finding such knowledge is important to finding the best balance between different policy instruments.

Hypothesis two: Existing knowledge structures (systems) are not adapted to crises and can lead to underutilization of knowledge

For over a decade, a lot has been written about the institutionalization of evaluation (e.g., Dahler-Larsen, 2019; Jacob et al., 2015; Leeuw & Furubo, 2008; Rist & Stame, 2006). Part of the institutionalization has been that special bodies have been set up to carry out evaluations, building evaluations systems, legislation, and clauses regarding what should be produced, at what intervals and how the information should be disseminated to decision-makers or others (Jacob et al., 2015; Leeuw & Furubo, 2008).

A similar institutionalization process has also taken place when it comes to other forms of knowledge production. Organizations like the Organisation for Economic Co-operation and Development (OECD) discuss the importance of science in constructing policies (e.g., Organisation for Economic Co-operation and Development, 2021). The European Commission has created specific bodies and structures to provide high quality, timely, and independent scientific advice for its policymaking activities. It seems that both nationally and internationally, it has been important to link expert structures to policy processes; their priorities are often based on assumptions about which knowledge they regard as important in the policy process.

In many countries, we can see how several research bodies, councils, and other organizational entities are linked to public decision-making through formal arrangements. The structure of knowledge production can be congruent with policy fields, while knowledge-producing bodies can be seen as part of sectorial structures such as education, employment, defense, environment, development assistance, and so on. In more stable situations, where the different steps in decision-making processes are well

planned, it is possible to overcome the problems caused by such a structure through different advisory and coordination processes.

The interface between decision-makers who ask for knowledge and policy advice (the demand side) and the knowledge producers (the supply side) can vary in different policy fields and situations. Most OECD countries seem to have several governmental expert bodies within economic and financial policies. In Sweden, there are several expert bodies with greater or lesser degrees of permanence that deliver direct, more immediate, policy-relevant analyses and studies in response to government requests for policy advice. The situation is similar in policy fields, as for example in education, the environment, and international development. And in case of a so-called Programme for International Student Assessment (PISA) shock, reflecting the OECD's assessment of student learning outcomes, which confronted several European countries, requests about policy advice can be addressed to several expert bodies representing different research fields and disciplines.

The degree of pluralism, overlap, or even competition, among expert structures, can certainly vary depending on the policy fields. However, and admitting that we do not have comparative studies about the governmental "knowledge system" in different policy fields and countries, it seems the interface between the policy side and the knowledge side has had distinct features during the pandemic compared with many other situations. In explaining this difference, two factors deserve to be further discussed.

The first factor has to do with the supply side. Due to the initial perception of this crisis, the degree of pluralism of engaged institutions was lower than in many other situations. With some exceptions, national public health institutes – which the International Association of National Public Health Institutes describes as a "government agency, or closely networked group of agencies, that provides science-based leadership, expertise, and coordination for a country's public health activities" (International Association of National Public Health Institutes, 2021) – took a very central and dominant position in delivering policy advice. In Sweden, the role of the national public health institute (Folkhälsomyndigheten), has been described as having a monopolistic position in the interpretation of existing knowledge (Jerneck, 2021, p. 13).

The second factor has to do with the demand side. The capacity to ask for policy advice and policy-relevant knowledge varies between different policy fields and policy questions. Although not surprisingly, it appears that political decision-makers are more accustomed to some policy questions, and some types of crises, than others. Questions related to, for example, economic policies, seemed to be on the agenda more frequently than those related directly to the pandemic. It is therefore possible that decision-makers are more accustomed to knowledge fields

and expert structures when they are facing urgent economic problems than to potential threats from a virus. Simply expressed, because the substantive issue was a virus, it was assumed, by the leading decision-makers, that the most relevant knowledge had to correspond with the decision-makers' initial perception. The search for policy advice was less open and pluralistic than in many other crises.

This factor can also be linked to the academic background of civil servants and political appointees within ministries who have a leading role in preparing policy decisions and in identifying which knowledge the political decision-makers need. It can be assumed that their backgrounds will impact what knowledge is sought and channeled into the political decision-making processes when a new problem arises. However, it has not been possible to access more detailed data which would have made it possible to compare the background among civil servants and political appointees connected with different policy fields. It can be noted that among civil servants and political appointees with a higher academic degree in Swedish Government Offices, economic and similar disciplines seem to be dominant, and the academic level seems to be higher in policy fields such as finance, employment, and economics.[6] At least in this hypothetical line of reasoning, I dare to assume that that civil servants with a more qualified background will more actively follow research within "their" field (e.g., through journals and networks), will have a broader intellectual understanding of conflicts and different perspectives among researchers within this field, and will therefore be likely to more actively search for contradictory information or at least information which can lead to awareness about different perspectives. The relatively homogenous background of civil servants and political appointees could have contributed to a limited capacity to seek out contradictory or at least additional information.

These factors can be assumed to have impacted the interface between decision-makers and the knowledge community. The initial perception of the crisis was decisive for which knowledge and knowledge bodies should deliver policy advice. Partly due to the academic background of civil servants and political appointees, it became difficult for final decision-makers to identify alternative approaches about which knowledge was policy-relevant.

Finally

The pandemic has been a crisis that has unfolded over a long time. Early interventions have been modified and sometimes abolished. This leads to a fundamental question for the evaluation community: to what degree has evaluation contributed to this process? Both the Swedish experiences and what is described in other chapters in this book such as those by Peter Wilkins and Ray Pawson seem to indicate that decisions

about different policy options have very seldom been made with reference to earlier evaluations. It can also be argued that references to concepts and tools that are central to "evaluative thinking" such as interventions theory, policy assumptions, side effects, and policy instruments do not seem to have been used very much in the political and administrative processes in which policies have been formed during the pandemic.

Whether this is surprising or not can be debated. In an earlier book in this series, it was argued that evaluation in crises had a very different role than in stable times, when decision-makers acted in the context of existing and ongoing interventions (Furubo et al., 2013). In crises, important decisions cannot be seen as a continuation of earlier policies and interventions. In a crisis, the important thing is not to find out how existing or earlier interventions worked or how observed problems can be solved through improvements in the existing intervention. When decision-makers are forced to find new courses of action, the value of evaluations conducted in the context of earlier interventions is limited.

Therefore, in crises, focusing on how earlier programs and interventions have worked is also not the most important source of knowledge. The questions raised can be described as much more ex-ante and open in character. This means that decision-makers need answers, based on the best possible knowledge, about the causal mechanisms to make it possible to verify or falsify assumptions on which different policy options are based.

Crisis therefore means that evaluation will have a more indirect relationship to decision-makers in stable times. When decision-makers ask for best possible knowledge, it is important that evaluation has contributed to building the basic knowledge about behavior and structures which Weiss talked about. When the findings of existing interventions are channeled into different disciplines and fields, the results from evaluation can be rearranged and discussed in relation to earlier data and theories. A discussion about the role of evaluation during the pandemic cannot therefore be only about the limited role it seems to have had during the pandemic. An important part of this discussion must also be about the extent evaluation, through its studies of different interventions, has contributed to a more fundamental knowledge that is useful in other situations when societies are facing new and urgent problems. It is obvious that it is much too early to discuss this question, but it should certainly be an important question in the future debate about evaluation and its role in society.

Notes

1 A broader discussion about the relation between crises, turbulence, and policy shifts is conducted in the introduction of an earlier book in this series (Furubo et al., 2013).

2 This statement can certainly be discussed further: it is based on a rationalistic approach that we are doing something because we will impact something else. However, some interventions can also be seen to have their own ends. For example, in some theories, punishment is its own end, because it restores the moral balance disturbed by crime (Ezorsky, 1972). And even in what we can consider rationalistic cultures, parts of the institutional framework and fundamental constitutional arrangements can be taken for granted and therefore are not an object for evaluations. An example can be taken from Sweden, where in common with a couple of other European countries, laymen have basically the same power in courts as professional judges. The system has developed over many centuries, and some aspects have been debated but it has never been evaluated (aside from minor questions within the system).
3 An early book in this series (Bemelmans-Videc et al., 1998) discusses different policy instruments and their use.
4 I will not in this chapter discuss the Swedish policy as such. However, it must be said that the Swedish government at least partly seems to have gradually changed its position and accepted more legally binding restrictions.
5 For an early overview of social and behavioral science relevant during the pandemic, see Van Bavel et al. (2020).
6 The statement is partly based on discussions with civil servants with extensive experience from different ministries. The material about the academic background among civil servants within Swedish ministries is limited. One study published a decade ago, pointed out the increasing number of economists on higher administrative levels in Sweden (Henrekson & Jakobsson, 2009). In preparing this chapter, I have requested figures from the administrative office within the government offices (förvaltningsavdelningen) about the academic discipline background among civil servants with higher education in different ministries. The answer I have received is that it is not possible to deliver such figures. However, it can be noted that the Swedish Government on the recruitment section of its website informs the reader that potential colleagues can be economists, lawyers, political scientists, public relations specialists, translators, IT specialists, and more (Regeringen, 2021). In this context, it is interesting that natural sciences including biology and medicine are not mentioned.

References

Bemelmans-Videc, M.-L., Rist, R. C., & Vedung, E. (Eds.). (1998). *Carrots, sticks and sermons: Policy instruments and their evaluation*. Transaction Publishers.

Boin, A., 't Hart, P., Stern, E., & Sundelius, B. (2005). *The politics of crisis management: Public leadership under pressure*. Cambridge University Press.

Boswell, C. (2009). *The political uses of expert knowledge*. Cambridge University Press.

Cederblad, J. (2020). *Munksydd eller inte? – frågan vägrar försvinna*. NyTeknik. www.nyteknik.se/samhalle/munskydd-eller-inte-fragan-vagrar-forsvinna-6996774

Chen, H.-T. (1990). *Theory driven evaluation*. Sage Publications.

Chen, H.-T. (2006). A theory-driven evaluation perspective on mixed methods research. *Research in the Schools*, *13*(1), 75–83.

Dahler-Larsen, P. (2019). The skeptical turn in evaluation. In J.-E. Furubo & N. Stame (Eds.), *The evaluation enterprise: A critical view* (pp. 144–163). Routledge.

Erlandsson, A. (2021). Smittminskningsprosocialitet. Hjälpandets psykologi i en pandemi. *Statsvetenskaplig tidskrift (The Swedish Journal of Political Science)*, *123*(5), 189–210.

Ezorsky, G. (1972). *Philosophical perspectives on punishment*. State University of New York Press.

Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Coupland, H., Mellan, T. A., Zhu, H., Berah, T., Eaton, J. W., Guzman, P. N. P., Schmit, N., Cilloni, L., Ainslie, K. E. C., Baguelin, M., Blake, I., Boonyasiri, A., Boyd, O., Cattarino, L., Ciavarella, C., … Bhatt, S. (2020). *Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries*. Imperial College. https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-03-30-COVID19-Report-13.pdf

Furubo, J.-E. (2019). Understanding the evaluation enterprise. In E. Furubo & N. Stame (Eds.), *The evaluation enterprise: A critical view* (pp. 3–31). Routledge.

Furubo, J.-E., Rist, R. C., & Speer, S. (Eds.). (2013). *Evaluation and turbulent times: Reflections on discipline in disarray*. Transaction Publishers.

Hedenvind, E. (2020, May 5). Anders Tegnell om munskydd: "Kunskapsunderlaget är extremt svagt". SVT Nyheter. www.svt.se/nyheter/inrikes/darfor-tror-inte-fhm-pa-munskydd-underlaget-ar-extremt-svagt

Henrekson, M., & Jakobsson, U. (2009). Vilket värde kan Institutet för Näringslivsforskning tillföra Svenskt Näringsliv och andra finansiärer? In M. Henrekson (Ed.), *IUI/IFN 1939–2009 Sju decennier av forskning om ett näringsliv i utveckling* (p. 434). Ekerlids förlag.

House, E. (1993). *Professional evaluation*. Sage.

International Association of National Public Health Institutes. (2021). *National public health institutes core functions and attributes*. https://www.ianphi.org/_includes/documents/sections/tools-resources/nphi-core-functions-and-attributes.pdf

Jacob, S., Speer, S., & Furubo, J.-E. (2015). The institutionalization of evaluation matters: Updating the international atlas of evaluation 10 years later. *Evaluation*, *21*(2), 6–31. https://doi.org/10.1177/1356389014564248

Jerneck, M. (2021). Inledning: Konsten att bemästra en pandemi. *Statsvetenskaplig Tidskrift*, *123*(2021/5). https://journals.lub.lu.se/st/article/view/23296/20712

Leeuw, F. L., & Furubo, J.-E. (2008). Evaluation systems: What are they and why study them. *Evaluation*, *14*(2), 157–169. https://doi.org/10.1177%2F1356389007087537

Leeuw, F. L., & Nelen, H. (2013). Evaluating counterterrorism: The case of the Netherlands. In J.-E. Furubo, R. C. Rist, & S. Speer (Eds.), *Evaluation and turbulent times: Reflections on a discipline in disarray* (pp. 197–228). Transaction Publishers.

Leonard, H. B., & Howitt, A. M. (2009). High performance in emergencies: Two modes of operation. In M. Howitt & B. Leonard (Eds.), *Managing crises: Response to large-scale emergencies* (pp. 611–624). CQ Press.

Merton, R. K. (1936). The unanticipated consequences of purposive social action. *American Sociological Review*, *1*(6), 894–904. http://www.jstor.org/stable/2084615?origin=JSTOR-pdf

Milstead, S. (2020). *Björn Olsen: Jag trodde inte på munskydden*. Aftonbladet. www.aftonbladet.se/nyheter/a/70EQR4/bjorn-olsen-jag-trodde-inte-pa-munskydden

Ministry of Social Affairs. (2020). *Äldreomsorgen under pandemin* (The elderly care in the pandemic), delbetänkande av Coronakommissionen (report of the Swedish Commission of Inquiry tasked with evaluating the measures to tackle the outbreak of the virus that causes the disease COVID-19). Regeringskansliet. https://www.regeringen.se/rattsliga-dokument/statens-offentliga-utredningar/2020/12/sou-202080/

Organisation for Economic Co-operation and Development. (2021). *OECD science and technology and innovation outlook 2021 – Times of crisis and opportunities*. www.oecd.org/sti/science-technology-innovation-outlook/crisis-and-opportunity/

Pawson, R., & Tilley, N. (2004). *Realist evaluation*. Design, Monitoring and Evaluation for Peace. https://www.dmeforpeace.org/wp-content/uploads/2017/06/RE_chapter.pdf

Pierson, P. (2004). *Politics in time: History, institutions, and social analysis*. Princeton University Press.

Regeringen. (2021). *Jobba hos oss*. Regeringskansliet. www.regeringen.se/jobba-hos-oss

Riecken, H. W. (1972). Memorandum on program evaluation. In C. Weiss (Ed.), *Evaluating actions program: Readings in social science and education* (pp. 85–104). Allyn and Bacon.

Rist, R. C., & Stame, N. (Eds.). (2006). *From studies to streams: Managing evaluative systems*. Transaction Publishers.

Rogers, P. J. (2008). Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation*, *14*(1), 29–48. https://doi-org.proxy3.library.mcgill.ca/10.1177%2F1356389007084674

Suchman, E. A. (1967). *Evaluative research: Principles and practice in public service and social action programs*. Russell Sage Foundation.

Van Bavel, J. J., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., Crockett, M. J., Crum, A. J., Douglas, K. M., Druckman, J. N., Drury, J., Dube, O., Ellemers, N., Finkel, E. J., Fowler, J. H., Gelfand, M., Han, S., Haslam, S. A., Jetten, J., … Willer, R. (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature*, *4*, 460–471. www.nature.com/articles/s41562-020-0884-z.pdf

Vedung, E. (2009). *Utvärdering i politik och förvaltning*. 3 uppl. Studentlitteratur.

Weiss, C. H. (1972). *Evaluation research: Methods of assessing program effectiveness*. Prentice Hall.

Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies* (2nd ed.). Prentice Hall.

World Health Organization. (2020a). *Strategic preparedness and response plan*. www.who.int/docs/default-source/coronaviruse/srp-04022020.pdf

World Health Organization. (2020b). *Covid-19 strategy update*. www.who.int/docs/default-source/coronaviruse/covid-strategy-update-14april2020.pdf

2 COVID Crisis

Time to Recalibrate Evaluation

Maria Barrados, Steve Montague, and Jim Blain

Introduction

When COVID-19 began to spread, few governments were prepared to deal with a pandemic (The Independent Panel for Pandemic Preparedness and Response, 2021). Governments were faced with dealing with a rapidly evolving crisis. Their decisions, made with limited information, little time to reflect, and requiring great urgency, have already led to many questions about the appropriateness of the responses to the COVID pandemic; these include questions about the costs and their effectiveness, particularly on the health care system's ability to weather the crisis and government measures to address the disruption, as the chapter by Maria Barrados and Jeremy Lonsdale in this volume illustrates. These questions will no doubt increase as governments face a subsequent period of recovery with demands for accountability and fulsome explanations. Indeed, auditors have already started the examination of financial and performance questions as can be seen in chapter by Barrados and Lonsdale.

What role have evaluators played in shaping the responses to the pandemic? Have evaluators built an agile and innovative enough practice to provide timely responses to questions about the results of the pandemic on existing programs and initiatives and how to assess the effectiveness of the pandemic response and recovery? Has the function helped with real-time learning, adjustment, and improvement? These questions are also explored in other chapters of this volume.

Evaluation offers approaches to knowledge production distinct from other disciplines and fields of study. Michael Quinn Patton (2020) has argued that "business as usual approaches" are inadequate to address fundamental transformations, which require an acknowledgment of the changes themselves and substantive knowledge of the challenges. Patton reiterates those arguments in his chapter in this book. Because of the disruption of the COVID pandemic, "business as usual" is not a viable option; rather, it presents an opportunity for improvement and recalibrating evaluation practices.

DOI: 10.4324/9781003376316-3

In a meta-analysis and website survey of evaluation key trends and learnings, Mariel Aramburu (2020) identified examples of such recalibration. For example:

- The World Bank Independent Evaluation Group and International Finance Corporation (IFC) discussed how to adjust rating processes and methodologies to account for shocks like the pandemic (proposals include rating projects based on their midcourse correction targets and giving the IFC more flexibility to choose the evaluation timing. It was suggested that this strategy may help projects recover and meet targets later (World Bank, 2020, p. 59).
- An analysis conducted by the United Nations' (UN) Office of Internal Oversight (OIOS) of 11 evaluation guidelines across the UN system, World Bank, and the Organisation for Economic Co-operation and Development (OECD) showed a consensus among all the entities that produced COVID-19 evaluation guidelines that it is no longer "business as usual" for evaluation either. The OIOS conclusion was that the function needs to repurpose and adapt both its focus and approach to reflect the limitations posed by the current crisis and ensure its utility in supporting the Organization's response (UN Evaluation Group, 2020, p. 8).

This chapter argues there is a great opportunity to recalibrate evaluation practice in response to ongoing pressures to demonstrate value and to respond to needs of users. Three innovative initiatives taken by Canadian evaluators are examined – one which supported management decisions on how to respond to the crisis, and two others, methodologically related, which show promise for recalibrating current practices to broaden the use and application of evaluation thinking and approaches, if used more widely.

Canada presents an interesting case study for a focus on the role of evaluation because the country has been viewed as one of the world leaders in evaluation culture. This positive perception is largely due to the institutionalization of the function and its past prominence in policy domains, at least at the federal government level (Jacobs et al., 2015). In 2016, the Canadian Government's "Policy on Evaluation" was replaced by a new "Policy on Results," setting out the requirements for performance information and evaluation, which still required evaluation, but gave it less policy prominence (Treasury Board Secretariat [TBS], 2016). This change reflected the pressure to provide more timely and ongoing information on effectiveness. As the policy required, most large federal departments proceeded to create distinct units for performance measurement, separate from evaluation and internal audit. Evaluation was already in large part co-managed with audit in most federal agencies.

Using examples from the Canadian federal government, three examples of innovation in evaluation are examined in this chapter to show where evaluation can use strengthen its practice. Evaluation can:

1. Play an active role in a crisis and proactively support management in dealing with difficult problems by examining immediate operational issues. An example is presented of how evaluation provided real-time support to senior management as they put in place the necessary functional structures to respond to the COVID-19 crisis;
2. take greater advantage of the potential synergies from co-location of evaluation and audit to provide deeper insight through exchange and collaborative work as illustrated by recent examples; and
3. act proactively to define what success looks like in the near, medium, and longer terms, thereby operationally defining and assessing key performance measures that help identify implementation issues that need to be addressed. An example shows how evaluation can provide an early warning system of program problems by identifying leading and lagging indicators.

Changing traditional practice to provide real-time management support

The challenge

One year into the COVID-19 pandemic, the Auditor General of Canada responded to a Canadian parliamentary request to examine the federal government's response. One of her reports concluded in essence that Canada had not been prepared for the pandemic (Office of the Auditor General [OAG], 2021). While recognizing the hard work of teams facing the crisis, the Auditor General's report is generally critical of the Public Health Agency of Canada (PHAC), part of the Health Portfolio, for not keeping emergency preparedness plans current and failing to ensure agreed improvements to surveillance and other functions were put in place. These and other gaps needed to be addressed quickly as the pandemic surged.

Evaluation approach

The PHAC faced a number of urgent operational challenges in managing the COVID-19 pandemic. In response to a request by the Chief Public Health Officer, PHAC's Office of Audit and Evaluation stepped up to help find workable solutions to deal with the crisis. They applied a methodology that provided a fast turnaround of practical help to senior management.

The PHAC evaluation team was actively involved in addressing early gaps – especially in terms of resourcing, capacity, and internal management

systems and support. In June 2020, the Office of Audit and Evaluation within PHAC took an "all-hands-on-deck" approach, which involved the following elements:

- Fifteen employees were assigned to the project – divided into various groups (e.g., interview group, document review group, report writers).
- Five themes for analysis were chosen among several potential topics which were determined to be the most critical areas to inform the organizational adjustments that could be made while the crisis continued.
- The themes (titles of each section) were determined by senior management before data collection and represented themes such as "Mobilization" and "Guidance."
- One evaluation lead and one manager were responsible for organizing the work. As reviews of various themes took place concurrently, smaller groups of evaluators collected and analyzed documents from public and internal sources for a document review within a specific theme. Document review took place over an intense two-week period.
- Generally, two or three people attended key informant interviews with key players. The interview guide applied to this work was characterized by being simple and "to the point." It resembled the structure of each section. One evaluator led the interview, a second took notes as close to verbatim as possible – then summarized them in a template which was grouped by theme. This allowed evaluators to do qualitative data analysis on the fly without using any special software, though they did further coding of notes in NVivo after the initial findings were prepared.
- The report was primarily drafted by three senior members of the evaluation group.

The project started in June and was completed in mid-July, with various presentations and reviews held throughout the rest of the summer. The report (with all five themes) was approved in September. Consultations with a team member suggest that the vast majority of data collection and reporting took place over the first three weeks.

The *Lessons Learned from the Public Health Agency of Canada COVID-19 Response (Phase One)* Final Report of September 2020 (Health Canada and Public Health Agency of Canada [PHAC], 2020) features a simple no-nonsense structure which limited background description and focused on:

- Skills, capacity, and mobilization;
- roles, responsibilities, and accountabilities for incident management and the various areas of work to address COVID-19;
- support to the Chief Public Health Officer;

- data to support decision-making; and
- guidance.

The structure within these sections typically laid out background (setting some context and describing in basic terms what was done); what worked well; challenges (generally identifying gaps or weaknesses); and suggested improvements. The report acknowledges its limitations in terms of its internal focus of consultation, but its rigorous documentation process allowed the group to get impressionistic – yet varied – inputs from key informants. The style of the report is reminiscent of a management letter in internal audit terms and clearly fits into a mode of communications which would be deemed "formative" in evaluation terms.

In addition to the specific suggested improvements in the five key areas (skills, capacity, and mobilization; roles, responsibilities, and accountabilities; support to the Chief Public Health Officer; data to inform decision-making; and guidance), the report also noted longstanding issues that had not been addressed such as confusion of roles with the Incident Management System (IMS). Suggestions were also made in cross-cutting areas.

Summary

As the report noted, work done by the PHAC and what was accomplished was unprecedented. There were notable successes, such as meeting diverse commitments under increasingly tight timelines and successful collaboration with partners, yet a lot still needed to be done. The Chief Public Health Officer and her management team had turned to the evaluation group to identify solutions to increase operational effectiveness and the quality of its response. Evaluation responded by implementing innovative methodology to provide timely and informed suggestions for improvement. The approach was considered so successful that at the time of writing it remains in use and efforts are underway to institutionalize this type of practice.

Using the potential of the co-location of evaluation and internal audit

Distinct practices with potential for sharing, collaborating, and learning

The PHAC case demonstrates how evaluation provided rapid, real-time input to support management in meeting the pressing operational requirements of dealing with the COVID-19 crisis in Canada. Two noteworthy features of the case are the co-location of audit and evaluation under one senior manager who was also the Chief Audit Executive and had trained as an evaluator. The second is a difference between the policy framework for evaluation and internal audit. The former emphasizes planning to evaluate

all the departmental expenditures and whether government requirements for evaluation are met, but it is silent about whether evaluation is encouraged to do other work. Internal audit standards, on the other hand, specifically allow internal audit to offer consulting services, which provides a policy rationale for the audit and evaluation group.

Consulting services are defined in the Internal Audit Standards as:

> Advisory and related client service activities, the nature and scope of which are agreed with the client, are intended to add value and improve an organization's governance, risk management, and control processes without the internal auditor assuming management responsibility. Examples include counsel, advice, facilitation, and training.
>
> (Institute of Internal Auditors, 2017, p. 22)

Audit and evaluation are distinct practices, which provide opportunities for the exchange of practices and greater collaboration between the two. As demonstrated in *Crossover of Audit and Evaluation Practices* (Barrados & Lonsdale, 2020), when audit and evaluation work together, they can gain from each other's expertise and both will have increased access to senior management time and exposure, as seen in the case of PHAC. However, as the book concludes, it takes special talent and leadership to manage these functions together. A deeper understanding of programs, opportunities for methodological innovation, and increased efficiencies for both audit and evaluation and the program itself can all be gained by working more closely together.

Internal audit and evaluation are required functions in Canadian Government departments and appear to provide considerable potential for sharing and collaboration. As argued by Lonsdale (2020, p. 208), academic research and public-sector practice has recognized that "the solutions to some of the greatest problems are unlikely to come from one discipline, and that individual disciplines or lines of inquiry can be significantly enhanced and far more effective by sharing of insights from elsewhere." In their recent book on audit and evaluation, Barrados and Lonsdale (2020) found examples of sharing and collaboration between performance audit practiced by external audit offices, evaluation, and evaluation and internal audit. Their review found examples where evaluation and internal audit worked together effectively and others where it proved to be more challenging.

The agility and responsiveness of evaluation in our case study, part of an Audit and Evaluation group responding to the COVID-19 crisis, demonstrate the potential for deeper collaboration and exchange between the two groups. Efforts to create greater synergy between these groups in Canada started in the 1990s with limited success. Internationally, the United Nations Joint Inspection Unit examined oversight functions within the UN system and recommended a consolidation in 2006 as a way to promote coordination and cooperation, avoid duplication, and create

synergy (Joint Inspection Unit [JIU], 2006). As described by Barrados and Lonsdale (2020), there have been different types of consolidation attempted and varied experiences as a result.

Internal audit and evaluation in Canada

Canadian Government departments, for the most part, have not capitalized on the organizational closeness of more mature evaluation and internal audit functions. A Canadian Treasury Board paper in 1993 examined linkages between internal audit and evaluation in Canadian federal departments in the late 1980s and early 1990s. The study found that most departments maintained separate audit and evaluation groups but had moved them organizationally closer together, most often reporting to the same senior manager. The motivation appears to have been primarily to save on executive resources and was not intended to have substantive effects on the work of either audit or evaluation. But as Hurd (1993) suggests, the common planning and budgeting resulted in a greater understanding of the two functions by the administrative head and the opportunity for each side to learn the methods of the other.

In Hurd's study of 15 of the larger departments, approximately half had completed a joint study with a further quarter of them open to the possibility in future. The authors found that "in most of these cases, however, the process of a joint venture resembles separate but coordinated audit and evaluation studies" rather than what would be considered a joint undertaking (Hurd, 1993). The author suggested that progress could not be made until the two were blended into a "Review" function (Hurd, 1993). This did not happen and many of the original structures of the 1980s and 1990s remain in place.

Many of the larger Canadian Government departments from the 1993 study have maintained the co-location of internal audit and evaluation almost three decades later. The most common structure is within corporate affairs, with separate heads of evaluation and internal audit with about a third of the larger departments having the same person as Chief Audit Executive and Head of Evaluation.

Again, from public reporting of internal audits and evaluations on the updated sample of 15 departments, there were few joint audits and evaluations found. As Frueh concluded in Barrados and Lonsdale (2020), her experience at UNESCO suggested that evaluation and audit collaboration can take different forms from formative sharing of methodological support and practice to completely joint studies (Frueh, 2020, p. 169). She also identified other collaboration such as more *ad hoc* approaches to integration reports later in the process and sequential reporting.

From public reporting of internal audits and evaluations in the updated sample of 15 departments, there were few joint audits and evaluations found. The above Health Canada example is a unique and formative

example. There was also an example of separate reporting – parallel audit and evaluation with separate reports on different aspects of the same topic, Jordan's Principle, which helped ensure First Nations children could access public support and services (Indigenous Services Canada, 2019a, 2019b). However, there are limited examples of joint work, which would appear to be the most challenging but also potentially have the most productive outcomes. Only two joint audit and evaluations from a line department and a central agency were found and will be explored.

Joint audit and evaluation (assessment) of NRCan's departmental governance, April 2021

A recent example of a joint audit and evaluation was conducted in the Department of Natural Resources of Canada (NRCan), which had not been part of the 1993 sample. This joint examination adhered to all the policies and standards for internal audit and evaluation. In its examination of the Department's governance arrangements, a joint approach facilitated the assessment of both internal controls and achievement of outcomes areas where both functions would traditionally have experience. As described in the report, the assessment was done against audit criteria that were derived from traditional audit concerns such as the presence of key controls and traditional evaluation concerns, such as expected results.

The resultant joint report concluded that "NRCan's governance arrangements are not sufficiently effective or efficient and are not providing the necessary value or relevant support to departmental oversight and decision-making" (Natural Resources Canada [NRCan], 2021). This conclusion was also expressed as an audit opinion. Both the conclusion and audit opinion were supported by a series of findings and recommendations. Management then provided an action plan in response to the recommendation.

Two noteworthy features of the report were, first, its clear labeling as a Joint Audit and Evaluation Assessment to differentiate it from either an internal audit or evaluation. Second, the report included a detailed appendix that combined the logic model from evaluation methodology with the audit risk map. Five sets of governance activities that lead to activities and immediate, intermediate, and ultimate outcomes were linked to audit risks and audit criteria (NRCan, 2021, Appendix C).

Reviewing the framing of this study, it is clear that there is a blend of what might be called "transactional" viewpoints (e.g., evidence of formal governance procedures and compliance to them) and "relational" viewpoints (e.g., demonstrated or self-reported understanding of roles and relationships and active participation in governance meetings). The observations and findings of the report are improved by this integration of viewpoints and approaches. Other organizations conducting a joint audit and evaluation for the first time have faced similar challenges in naming

the resultant report and developing a common understanding of the different use of common terms and procedures (Barrados & Lonsdale, 2020, p. 183). The NRCan report reflects its approach to the same challenges by adding "Assessment" to the title of the report and including an appendix showing the linkages between the audit and evaluation concepts. The report is clear, however, that both audit and evaluation met the required policies and standards.

Joint audit and evaluation of privacy practices at the Treasury Board of Canada Secretariat

In one report, evaluation and internal audit examined the performance, practices, and controls of the privacy practices at the Treasury Board Secretariat (TBS), the body supporting the federal government management board. By working together, the team reduced duplication and burden on the small group being examined. Interviews were conducted jointly but auditors and evaluators also undertook independent analysis before jointly reaching conclusions and recommendations.

The report presented three sets of findings related to awareness and understanding, privacy and assessment tools, and resources and made five recommendations, as well as including a management response and action plan.

In a presentation to the Canadian Evaluation Society, the team leads of the project Natalie Lalonde and Elena Petrus (2021) described their experience. They had searched the literature which did not provide them with any guidance. There were a number of practical issues, such as where evidence and data would be held and the timing of different steps in the audit or evaluation process that were easily resolved. However, their biggest challenge came when they realized that terminology, such as "root causes" and even "findings" used by each discipline did not have a common understanding. This made them realize that even though they had been working within the same organizational structure and even worked on joint planning, they did not fully understand each other's practice.

The TBS audit and evaluation team had one member who had worked both as an auditor and evaluator. This person pulled the material together in a single report that met the standards for both internal audit and evaluation. The report included a graphic illustrating the relationships between the evaluation questions, audit criteria, conclusions, and logic model in its appendices (TBS, 2021). The figure illustrates which components and elements of the Treasury Board of Canada Secretariat Access to Information and Privacy Office's logic model were addressed in the evaluation and which in the audit. The figure illustrates that evaluation had questions and audit had criteria. Both intermediate and immediate outcomes were addressed jointly, with some relying more on evaluative results. Longer term outcome questions and conclusions were dealt with by the evaluators.

The evaluation form of inquiry involved an open-ended set of questions on how or what was being done, which evaluators assessed against program and policy objectives. By contrast, auditors had a framework defining which management processes and controls were to be in place and drew their observations and conclusions against them.

Both approaches contributed to the final conclusions and recommendations. The project leaders concluded that their joint work resulted in developing a greater mutual understanding of the contributions of each practice (Lalonde & Petrus, 2021). With limited resources and many questions about program and policy performance, the joint work provided efficiencies in delivering the project for both the practitioners and those being assessed with a more useful, comprehensive output which brought distinctive perspectives together to produce improved insight.

Unrealized opportunities to gain efficiencies and report usefulness

The early promise that closer organizational location would improve the mutual understanding of the audit and evaluation practices appears to have stalled primarily at the budgeting and macro planning level. The examples of the two joint audit and evaluations illustrate that there is value in a combined study, but also that time and effort was required to facilitate joint working so that team members were comfortable with their respective terminologies and methodologies.

Collaboration between internal audit and evaluation work does not necessarily imply following a single approach, nor does it necessarily result in a single report. Closer collaboration could mean using a more operational, consultative approach familiar to internal audit together with evaluation's strength in data gathering techniques to address the question of how to deal with a crisis, such as in Health Canada's response to COVID-19. Frueh concluded that the approach to audit and evaluation work should not be standardized but depend on the context and the key issues to be addressed (Frueh, 2020, p. 181).

There will no doubt be challenges, but with strong leadership from the management of both functions, mostly co-located but with some joint management, there can be further exchange of practices and methodologies, as well as an increase in joint audit-evaluation projects. Leaders will need to have a clear, constructive, and unbiased approach that favors neither function but champions both to deal with inherent differences in policies and procedures. Government evaluators and internal auditors can look to the practice of performance audit in national audit offices, where many evaluators have had long careers using evaluation methodology and techniques within an external audit office with performance audit standards. The performance audit standards provide the necessary flexibilities to do "evaluative work" using evaluation methodologies. However, this does require evaluation team members to take training on the standards.

The opportunity exists for evaluators, as part of macro budgeting and planning to also identify areas that could be strengthened through collaboration with internal audit. Collaboration with evaluation is an opportunity to increase efficiency and productivity of internal audit; for evaluators, it offers greater scope and opportunity to demonstrate the value of their work.

Developing leading indicators

Throughout the evolving COVID crisis, there were many questions about what could work and what would not. For example, there were early questions about the efficacy of wearing masks, as well as the best ways to support citizens as parts of the economy shut down. There were also questions about what types of messaging and channels were most effective to influence citizen behaviors regarding mask wearing, vaccinations, and maintaining physical distancing. As illustrated in the PHAC case, information to support this kind of decision-making is not always at hand. The hope from the long-standing effort at evaluation and the new emphasis on performance information is that more information in the form of leading and lagging indicators would be available to better target interventions – and to do so in real time.

Administrative data alone tends to be insufficient. As Heinrich (2002) concluded in the case of US federal job-training programs, results of empirical analysis confirm that the use of stand-alone administrative data on performance management, in itself, is unlikely to produce direct, accurate measures of fully comprehensive program impacts.

In many respects, performance measurement and evaluation are complementary activities with a shared purpose of providing information on results. The development of performance information relies on evaluative thinking and methodology (McDavid et al., 2018). Program evaluations perform in-depth studies, while performance measurement provides the results information which managers need on programs. Hatry (2013), in his examination of evaluation and performance measurement, concluded that one area for improvement was for public officials to better estimate the future costs and outcomes of initiatives. During a time of crisis, the need for predictive information based on rigorous analysis is particularly important.

Evaluation practices, which may have been locked into a planned cycle of studies oriented to summative judgments, also have the opportunity to identify predictive indicators that support program managers as part of these studies and, in a time of crisis, provide robust information to support decision-making. Evaluators can expand the use of the traditional spectrum or logical chain of investment, activities, outputs and outcomes, concepts, and indicators to identify predictive indicators and measures. Rigorous thinking on how a program works can help managers in a practical way by developing and using predictive measures to better support management decisions in operational as well as tactical and

strategic management. These measures would also be ideal candidates for performance measurement results frameworks used for management and accountability. For evaluation, robust data collection and analysis would supplement evaluation findings with predictive performance indicators.

Evaluation and predictive indicators

There are several contexts in which predictive or lead indicator concepts have evolved, sometimes referred to as an "early warning system." Whatever terminology is used, the approach is expected to have significant predictive power with respect to future projected performance impacts and outcomes (ex-ante orientation). They would be expected to provide insight into the factors that drive results and how they can be measured and monitored.

Perhaps the best-known lead indicator application deals with the entire economy, usually updated quarterly or annually. Examples of lead indicators of future economic growth performance are stock market performance, manufacturing activity levels, inventory levels, retail sales, building permits, housing statistics, and new business start-ups. All these aggregate indicators were closely followed as the COVID crisis evolved. These macro indicators are not based on strong theory but are the result of observations over time.

The challenge for evaluation and performance measurement is to support governments by providing timely and reliable information that provides the facts to respond to this complex, rapidly changing public service environment, especially during unexpected events like the COVID pandemic. Evaluation through in-depth analysis can provide valuable insight into how results are achieved. Some analysis could further be used to provide program management with individual performance measures that could support better management of their programs. Information, computing, and communications technology improvements bring with them the opportunity to harvest more data faster (Petersson & Breul, 2017).

Most evaluations, compared to performance measurement regimes, typically include the concept of making a judgment of value or worth and of cause and effect. They traditionally have been considered an assessment of a program, policy, or activity once it has been completed (ex-post). The analysis of cause and effect would therefore help to identify factors that contributed to its success. If circumstances did not significantly change, these factors could also be predictive of future success.

The opportunity exists to view the ex-post evaluation as the source of indicators that can be considered as a predictive indicator in the same way as economists use leading indicators. The analysis of the illustrative example below was previously published; however, the logic and application of the case remains current and presents an ongoing gap for evaluation.

Predictive indicators of why staffing was so slow

In April 2000, the Auditor General of Canada reported to the Canadian Parliament that there was an urgent need to address long-standing problems in human resource (HR) management, with staffing in particular identified as a major source of frustration (OAG, 2000, 9.97). Staffing was viewed as unduly complex, inflexible, and inefficient, taking twice as long as in other quasi-public organizations. Its reform was described by the Auditor General as "imperative."

Numerous studies, including those by the Public Service Commission, which is responsible for public service appointments, recommended fundamental changes to the staffing system and its legislative framework. These studies recommended changes in practice that required revisions to enabling legislation. In response to concerns about the inefficiency in the process, Parliament passed the *Public Service Modernization Act* (PSMA) in 2003. It was described as the most significant legislative change in public service human resource management in the past 35 years. Despite the changes in legislation and policies, delegations by the Public Service Commission, and new training for managers, the problem of slow staffing processes persisted. The challenge was how to encourage managers and staff of departments and agencies to further change their practice. Data was available from analytic studies, audit, databases, and results from the staffing management accountability agreement. An ex-post evaluation approach was used to determine under which circumstances the new legislative and policy framework was successful in reducing the time to hire staff. An analysis of single staffing processes for the permanent positions filled within government for 46 federal departments and agencies was used. These organizations accounted for 90 percent of staffing activity (Barrados & Blain, 2013).

The analysis found that the key factor which linked to staffing efficiency (i.e., reducing the time to recruit) was a "robust results-based management accountability system (which includes an HR planning component)." This factor was defined as a system whereby managers compare results to plans and adjust their HR plans to address identified problems. Neither planning on its own nor increasing front-line HR staff was found to produce increased efficiencies.

The Commission concluded that:

> When organizations were focused on accountability and assessed HR results against concrete, realistic plans and strategies … the times to staff were reduced by as much as 30 days compared to other organizations.
>
> (Public Service Commission [PSC], 2009, 5.62)

These results argue for using these predictive operational measures to increase staffing efficiency and as guides to make necessary management changes. They also serve as an "early warning system" of impending

poorer results. Combined with periodic evaluation, the use of a predictive lead indicator has greater prospects of guiding management action to improve performance. By using a predictive indicator derived from the analysis of the logic chain to achieve a specific result with ex-post evaluation, managers are provided with a tool that will help improve their practice to achieve a better result.

The approach worked because of the existing underlying management framework supported by legislation. The evaluators worked closely with policy analysts who monitored program performance. The empirical test confirms the measure of predictability that provides for more rapid, ongoing adjustment and direction than a cyclical evaluation. By paying attention to this one particular element, managers could significantly improve results in an area of mutual concern.

Leading indicators can be set up to suggest the key behaviors and actions which predict success. In the world of staffing, organizations such as governments should have reasonably good access to internally generated information on the degree to which agencies have implemented robust accountability systems which include an HR component. The example illustrates that groups which took HR planning and variance management seriously, evidenced by actively comparing results to plans and then "managing for results," showed reduced staffing times.

The value added of evaluation identifying predictive indicators

Identifying predictive indicators depends on having a rigorous understanding of how a program or intervention works. This can be achieved either by drawing on existing evaluations and other analytic work, as the staffing example demonstrated, or by taking the time to lay out the logic of the program in relationship to other explanatory frameworks and then identifying patterns which match similar types of interventions.

The identification of predictive measures through evaluation provides for an analytic rationale for target setting and a means of doing more ongoing assessments of effectiveness rather than relying on periodic evaluations. Furthermore, it establishes a rich source of information available to help address immediate problems and crisis situations such as the COVID pandemic. This supports the argument made by Wholey (2010) for a closer integration of performance measures and evaluation. An integrated approach to program evaluation and predictive analytics can be mutually beneficial. Such an approach provides ongoing contextual information to evaluation and an analytic inferential base for performance measurement.

This integrated approach is particularly important in view of recent progress in predictive analytics as a result of improvements in the design and coverage of administrative data systems and improved access to such data systems (Davenport & Jarvenpaa, 2008).

Once they are integrated with program evaluations, individual predictive operational performance indicators, because of their ongoing nature, can also serve as an action-oriented "early warning system" for program management, and therefore enable a more rapid response to problem areas as they emerge. This circumvents the need to wait for a periodic five-year evaluation to be undertaken in future.

Corrective interventions occurring sooner rather than later provide for more timely adaptation and improved efficiency in bureaucratic organizations. Ongoing performance measurement systems can be strategically designed to include a series of predictive bench-mark indicators, along with a range of short-term feedback information in areas linked to future, long-term success. Such predictors can legitimately be regarded as "lead indicators" – if the problems they identify or confirm are left untreated in the short-term, there can be a cumulative effect by the time the program is subjected to a longer term, periodic retrospective in-depth evaluation.

For the purposes of government decision-making and change management initiatives, opportunities exist to create stronger synergies between evaluation and performance measurement to provide more timely and forward-looking information for government decision makers and better accountability.

Conclusion

From our observations and publicly available information, we found one clear example where evaluation played a "real-time" role in the Canadian Government's response to COVID-19 by providing information useful to decisions on how the Government should respond. The evaluators in this case were agile and innovative but also maintained a rigorous approach to provide the results of their work in a timely way.

The information and advice gathered was operationally focused and formative rather than summative. The work of the evaluators benefited from being able to draw on internal audit practices and presence at the management table. The work that was done by the PHAC group was in the context of the urgent need for information at the time of the crisis.

Evaluators are accustomed to the challenges of doing different types of evaluations – from project evaluation to operational evaluations, formative evaluations, impact evaluations and strategic evaluations. The latter areas are typically the most time-consuming, requiring the most in-depth analysis, and usually addressing a specific request. It is in the area of project, operational, and formative evaluations that opportunities exist for greater innovation and agility on the part of evaluators that would provide more timely and fulsome responses that are needed for day-to-day management decisions in a time of a crisis.

In addition to the example of the work at the PHAC, two other initiatives were identified where evaluation could help recalibrate some

practices. The other initiatives show the potential and benefits of greater collaboration with other established groups in the Canadian federal Government – internal audit and performance measurement. As is the case for internal audit, there is sufficient flexibility in evaluation standards to provide for collaborative work, with distinct practices from internal audit. Internal audit tends to examine control, economy, and efficiency issues while evaluation focuses on the contribution of initiatives to outcomes and effectiveness. Together, they can present a more complete picture of program operations. As the examples illustrate, there are challenges in collaboration, but they can be overcome with leadership and a commitment to understanding each other's work.

The linkage with evaluation and performance measurement is a well-established part of the methodological tool kit of evaluators. In the past, Canadian evaluation units had been formally required to identify indicators for new programs as a means to assess their future performance. There has been less emphasis on the role that evaluation can play in identifying predictive indicators, but recent cases suggest that this is one area that has potential promise.

Evaluation can establish research-based patterns of linked results (such as established theories of change and results pathways). It can focus on early indicators of delivery, reach, engagement, reaction, and relationships and under which enabling conditions can link to desired results and sustained impacts. This amounts to systematically seeking to develop, test, and refine leading indicators which can predict later outcomes. This in turn can systematically establish what works, to what extent, for whom, under which conditions, and why.

The disruption of the COVID-19 pandemic means "business as usual" is not a viable option. Rather, it presents an opportunity for improving and recalibrating evaluation practices. The recalibration of evaluation to link with other functions and to focus on real-time management support, such as in the cases described, seems to be in order. Such changes will have significant policy, planning, strategy, resourcing, and structural implications that can also have significant benefits for the evaluation function, strengthening its contribution at the management table.

References

Aramburu, M. (2020). Learnings in relation to COVID-19. Unpublished working paper.

Barrados, M., & Blain, J. S. (2013). Improving program results through the use of predictive operational performance indicators: A Canadian case study. *American Journal of Evaluation*, *34*(1), 45–56. https://doi.org/10.1177/1098214012464426

Barrados, M., & Lonsdale, J. (Eds.). (2020). *Crossover of audit and evaluation practices: Challenges and opportunities*. Routledge.

Davenport, T. H., & Jarvenpaa, S. L. (2008). *Strategic use of analytics in government*. IBM Center for Business of Government. https://www.businessofgovernment.org/sites/default/files/Strategic%20Analytics.pdf

Frueh, S. (2020). Two sides of the same coin: The UNESCO example. In M. Barrados & J. Lonsdale (Eds.), *Crossover of audit and evaluation practices: Challenges and opportunities* (pp. 161–181). Routledge.

Hatry, H. P. (2013). Sorting the relationships among performance measurement, program evaluation and performance management. *New Directions for Evaluation*, *2013*(137), 19–32. https://doi.org/10.1002/ev.20043

Health Canada and Public Health Agency of Canada. (2020). *Lessons learned from the Public Health Agency of Canada's COVID-19 response (phase one) final report*. House of Commons. https://ourcommons.azureedge.net/data/Motion2020-10-26/2021-01-20-SPDP-8550-432-2/HESA_PHAC-ASPC_0000186_E.pdf

Heinrich, C. J. (2002). Outcome-based performance management in the public sector: Implications for government accountability and effectiveness. *Public Administration Review*, *62*(6), 712–726. https://doi-org.proxy3.library.mcgill.ca/10.1111/1540-6210.00253

Hurd, I. (1993). *Linkages between audit and evaluation in Canadian federal departments*. Treasury Board Secretariat. https://www.tbs-sct.gc.ca/pubs_pol/dcgpubs/tb_h4/evaluationtb-eng.asp

Indigenous Services Canada. (2019a). *A review of Jordan's Principle*. https://www.sac-isc.gc.ca/eng/1565786653117/1565786690869

Indigenous Services Canada. (2019b). *Audit of the implantation of Jordan's Principle*. Retrieved from https://www.sac-isc.gc.ca/eng/1594378735468/1594378764255

Institute of Internal Auditors. (2017). *The International Standards for the Professional Practice of Internal Auditing (Standards)*. https://www.theiia.org/en/content/guidance/mandatory/standards/international-standards-for-the-professional-practice-of-internal-auditing/

Jacobs, S., Speer, S., & Furbo, J.-E. (Eds.). (2015). The institutionalization of evaluation matters: Updating the international atlas of evaluation 10 years later. *Evaluation*, *2*(1) 6–31. https://doi-org.proxy3.library.mcgill.ca/10.1177%2F1356389014564248

Joint Inspection Unit. (2006). *Oversight lacunae in the United Nations system*. https://www.unjiu.org/sites/www.unjiu.org/files/jiu_document_files/products/en/reports-notes/JIU%20Products/JIU_REP_2006_2_English.pdf

Lalonde, N. M., & Petrus, E. (2021). *Audit and evaluation: Better together?* [Presentation]. Canadian Evaluation Conference 2021, Ottawa, ON, Canada.

Lonsdale, J. (2020). Conclusion. In M. Barrados & J. Lonsdale (Eds.), *Crossover of audit and evaluation practices: Challenges and opportunities* (pp. 208–219). Routledge.

McDavid, J. C., Huse, I., & Hawthorn, L. R. L. (2018). *Program evaluation and performance measurement: An introduction to practice* (3rd ed.). Sage Publications.

National Resources Canada. (2021). *Joint audit and evaluation (assessment) of NRCan's departmental governance*. https://www.nrcan.gc.ca/transparency/reporting-and-accountability/plans-and-performance-reports/audit-and-evaluation/reports-year/joint-audit-and-evaluation-assessment-nrcans-departmental-governance/23631

Office of the Auditor General of Canada. (2000). *Report of the Auditor General of Canada. Chapter 9: Streamlining the human resource management regime – A study of changing roles and responsibilities*. https://publications.gc.ca/site/fra/409299/publication.html

Office of the Auditor General of Canada. (2021). *Report 8 – Pandemic preparedness, surveillance, and border control measures*. https://www.oag-bvg.gc.ca/internet/English/parl_oag_202103_03_e_43785.html

Patton, M. Q. (2020, March 27). *Evaluation implications of the coronavirus emergency*. Social Science Space. https://www.socialsciencespace.com/2020/03/evaluation-implications-of-the-coronavirus-emergency/

Petersson, G. J., & Breul, J. D. (Eds.). (2017). *Cyber society, big data and evaluation*. Transaction Publisher.

Public Service Commission of Canada. (2009). *Public Service Commission 2008–2009 annual report*. https://publications.gc.ca/site/eng/9.671529/publication.html

The Independent Panel for Pandemic Preparedness and Response. (2021, May). *How an outbreak became a pandemic: The defining moments of the COVID-19 pandemic.* https://theindependentpanel.org/wp-content/uploads/2021/05/How-an-outbreak-became-a-pandemic_final.pdf

Treasury Board. (2016). *Policy on results*. Government of Canada. https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=31300

Treasury Board. (2021). *Joint Audit and Evaluation of Privacy Practices at Treasury Board Secretariat*. Government of Canada. https://www.canada.ca/en/treasury-board-secretariat/corporate/transparency/internal-audit-evaluation-treasury-board-canada-secretariat/joint-audit-evaluation-privacy-practices-tbs.html

United Nations Evaluation Group. (2020). *Synthesis of guidelines for UN evaluation under COVID-19*. http://www.uneval.org/document/detail/2863

Wholey, J. S. (2010). Overcoming challenges to performance measurement and evaluation. *The Public Manager, 39*(4), 63–69.

World Bank. (2020). *Results and performance of the World Bank Group 2020: An independent evaluation*. http://hdl.handle.net/10986/34844

3 The Unbearable Lightness of Rights

Evaluation and COVID-19 Responses

Pearl Eliadis[1]

Introduction

The World Health Organization declared a global pandemic on March 11, 2020, in response to the threat of the novel coronavirus, COVID-19. The declaration triggered a scramble for data to assess disease progress, using a wide range of health and social measures to contain the spread and prevent disease resurgence (Prevent Epidemics, 2020). By mid-2022, some estimates placed the number of deaths at about 15 million people. While this figure represents a fraction of the deaths recorded during the 1918 Spanish flu, which killed more than 50 million people, the current pandemic has been catastrophic at many levels. Greater travel and mobility, high levels of contagion (especially for the Omicron variant), and the relatively recent capacity of modern economies to provide health and social assistance have all meant that the social and economic consequences have been devastating in ways that would have been impossible a century ago. A widely cited paper by David Cutler and Lawrence Summers (2020) estimated that even in the early days of the pandemic in 2020, the COVID-19 virus had already cost the United States alone $16 trillion, while the global cost, estimated by extension from the US economy, was in the region of $96 trillion.

While recognizing the severity of the socioeconomic implications, there has been a growing awareness of the extent to which much of the analysis has been disconnected from holistic, rights-based approaches, even within the healthcare sector itself (Gianella et al., 2020). The consequences of this disconnect are deeply worrying, especially for developing economies that are likely to experience the most lasting and long-term damage (Yeyati & Filippini, 2021). While many public policies recognize that vulnerable and marginalized people should have received priority attention, the reality during the pandemic suggested the opposite and veered in the wrong direction. Older persons (Landry et al., 2020), ethnic communities and women (Connor et al., 2020; Gianella et al., 2020), and people with disabilities (Negrini et al., 2020), to a name but a few, have been grievously and disproportionately affected by the pandemic. Existing social fault lines

DOI: 10.4324/9781003376316-4

were exacerbated while new ones yawned open, exposing systemic and structural barriers and discrimination for people whose basic rights were already precarious.

Evaluators should have been first in line to both notice and address the magnitude of the human rights issues that were appearing on the horizon. Human rights should have had pride of place in efforts to respond and to "build back better" (Kjaerum et al., 2021). Michael Quinn Patton (2020) has called for nothing less than a substantial systems transformation. His call to action is reiterated in his chapter in this volume. And yet, as this chapter argues, human rights law and human rights-based approaches (HRBAs) have not been centered within evaluation practice, even in the post-crisis era. There is currently no evidence that the status quo is likely to change, as evaluators continue to focus heavily on methodological issues and shy away from substantive issues like human rights that may raise controversy.

This chapter begins by examining why that might be, starting with an inquiry into the weak role that human rights play in practice within most of evaluation's ethical frameworks. It also argues for a substantive strengthening of both human rights and HRBA as the central load-bearing beams in evaluation practice.

The second section sets out the case for why human rights matter, not just because of legality or lawfulness, but because they form part of the global order, starting with the United Nations Charter. They are essential for ensuring the rule of law and the coherence and legitimacy of public policy. Consequently, it is argued that human rights should be front-loaded transparently in evaluation practice, despite the acknowledged challenges posed by human rights. An example of the challenge – and the imperative – of integrating HRBA into to evaluation in economic and social policy areas is discussed through the example of the right to adequate housing. The right to adequate housing, as a human right, is rapidly emerging in many countries as a central policy plank as it becomes clearer that the private market model has completely failed to generate adequate supplies of accessible and affordable housing in developed countries like Canada. Housing markets have heated up, and unaffordability has pushed people into homelessness or into inadequate housing that threatens individual lives, health, and communities, with knock-on effects for virtually all human rights. These effects have been magnified during the pandemic and exacerbated by diminished access to services and disrupted supply chains. HRBAs have made a difference to government efforts to tackle this complex problem in Finland and Scotland, among others, and have started to change the policy approach in countries like Canada.

The third section of the chapter is devoted to the role of human rights and HRBA in times of emergency so that evaluators can better understand the criteria that should be used, especially measures that prioritize human

life. It reviews how and when public measures can legally derogate from human rights during times of emergency and discusses the role of non-discrimination as a key tool in assessing emergency measures.

Evaluation, ethics, and human rights: The vanishing line

There are many reasons why evaluators may have reacted slowly during the pandemic and why human rights have not been centered in evaluation responses. In his chapter in this book, Jan-Eric Furubo explains that such reactions are typical across all fields: emergencies are critical junctures or breaking points that render decision-makers reluctant or unwilling to acknowledge the scope, scale, and significance of the events triggered by a major crisis. Focusing more specifically on evaluation, Indran Naidoo's chapter in this volume describes the United Nations' evaluation system's slow response to COVID-19, where evaluators were rapidly overtaken by other knowledge providers who asserted their influence with donors and policy makers. The issue is not only lack of rapid reaction but also lack of a reaction in favor of a strong orientation toward human rights, a phenomenon that is all the more surprising in light of the 2020 *Call to Action for Human Rights*, positioned as "the highest aspiration" aimed at "a human rights vision that is transformative" (UN, 2020). That vision should have been front and center for evaluators.

Clearly, these responses are not the product of a lack of value statements, ethical frameworks, or normative criteria. Not only should evaluators do no harm, we are told, but they should also "do good" and "tackle bad" (van den Berg et al., 2022). They should also establish what good "ought to be" (Stame, 2018). National and international bodies have developed detailed guidance on ethical practices to unpack what these general injunctions mean. The guidance includes the Joint Committee Standards for Educational Evaluation's Program Evaluation Standards (Yarbrough et al., 2010), the American Evaluation Association's (AEA) Guiding Principles (American Evaluation Association [AEA], 2018), and the Organisation for Economic Co-operation and Development's (OECD) – Development Assistance Commission (DAC) Network on Development Evaluation *Summary of Key Norms and Standards* (2010). These various standards set out evaluation criteria such as accountability, effectiveness, efficiency, feasibility, impact, propriety, relevance, sustainability, and utility. However, they show weak connections to human rights law and HRBA.

The Joint Committee Standards, for example refer to "human rights and respect" as a mere sub-category of "propriety standards." "Propriety" connotes conventionally accepted standards of behavior or morals but conveys nothing of the legal obligations inherent in international human rights law and its instantiation in domestic constitutions and laws. The Joint Committee Standards simply posit that legal rights and human rights comprise but one

of several important considerations, such as "clarity," or "responsive and inclusive orientations" toward stakeholders and targeted communities.

Turning to the AEA *Guiding Principles*, a principle of the "common good and equity" is proposed, but the *Guiding Principles* lack any discussion of equality, access to justice, gender equality, or even human rights more generally. Finally, the OECD DAC *Summary of Key Norms and Standards* (2010) contains 34 pages of principles, criteria, standards, and evaluation guidance, in which human rights appear exactly twice. The first appearance is contained in a section called "evaluation ethics" that addresses respect for "human rights and differences in culture, customs, religious beliefs and practices of all stakeholders" (p. 20). The second is as a type of "crosscutting issue" like gender or the environment (p. 23). All the elements appear to be given similar weight with the result that the normative force of human rights is diluted.

This framing and the resulting dilution of human rights' intended influence can foster the very conditions in which those rights are likely to be minimized and ignored. Worse, some of the stated ethical standards may operate contrary to human rights law. Respect for cultures, customs, and practices, for example, is an important consideration, but it may also enable harmful customs and practices that violate human rights, especially those of migrants, minorities, women, and children.[2] Professionals engaged in evaluation work in the development context, including this author, are frequently told not to "interfere," or are encouraged to engage in relativist arguments. It bears mentioning at this point that it is precisely this sort of thinking that forced the human rights community to insist that women's rights are human rights, because institutions did not want to get involved in "private" matters. The same applies to the relativist thinking that required a reaffirmation of human rights as universal and interdependent in the 1993 *Vienna Declaration and Programme of Action*. Violations of human rights related to rights such as the right to life, liberty, and security of person, among many others, give rise to obligations of immediate realization and engage with clear legal obligations. While not all rights fall into this category and while States have leeway in subjecting certain rights to reasonable limits, an analysis is needed to distinguish among rights. That analysis is often sorely lacking.

Rob D. van den Berg examined ethical standards from 11 major evaluation sources and guidance from national and international evaluation societies and groups; 44 terms are used in the selected evaluation frameworks that have ethical connotations of which human rights is but one (2022). Alarmingly, van den Berg found that human rights and lawfulness did not score significantly better than many of the 44 ethical standards in terms of prevalence.

Examining these ethical frameworks and insights into the types of ethical injunctions, one cannot help but be underwhelmed. Human rights should not be one of many similar "do good," "it-would-be-good," or

"at least do no harm" criteria. The above standards and principles offer little in the way of human rights content and lack any organizing principle to distinguish among them or decide which standards matter more. Few of them differentiate optional norms and those with legal force, let alone those standards that are grounded in the international human rights framework. Human rights are not just any kind of law: they are one of the three pillars of the United Nations (United Nations, 1945). Human rights are protected by international law, national constitutions, and human rights legislation. "Propriety," "professionalism," and "do no harm" and the like, lack the normative force of human rights law.

After the pandemic began to subside and when measures began to lift, matters did not improve much in terms of the emphasis placed on human rights in evaluation standards. Many of the new standards and criteria that emerged, failed to venture much beyond recycled restatements of existing evaluation standards or tweaks to well-known criteria. As part of the research undertaken for this book, a meta-analysis and web survey was conducted for evaluation publications in English and in Spanish in the first year of the pandemic (March 2020–February 2021).[3] The organizations reviewed were:

- Four multilateral development banks[4];
- four sample UN entities[5];
- eight regional evaluation societies[6];
- eight country-level evaluation societies[7];
- the International Development Evaluation Association (IDEAS); and
- the OECD (DaC EvalNet)/United Nations Development Programme.

Few of the publications from these organizations contain human rights standards except those that were embedded generically in ethics statements like those discussed earlier in this section. The United Nations Evaluation Group's Synthesis of Guidelines for UN Evaluation Under Covid-19 (Office of Internal Oversight Services, 2020) reviewed 11 guidelines from the UN system but makes no reference to human rights in its common themes and over-arching dimensions although the UN Women's Pocket Tool for Managing Evaluation during the COVID-19 Pandemic refers to HRBA (United Nations, Women, 2022).

Moving from human rights law to HRBA, and despite what appears to be an obvious area of focus for evaluators, it is relatively rare to see HRBA informing evaluation theory and practice outside of human rights law and rule of law projects. In addition to the weakness of human rights in evaluation standards and criteria that have already been discussed, none of the more than 30 volumes in the *Comparative Policy Evaluation* series, of which this volume forms a part, centers human rights as an organizing principle for evaluation practice, although one publication refers broadly to social justice with a focus on equity (Forss & Marra, 2014).

In conclusion, and except for specialized pockets of evaluation practice, human rights possess an undistinguished place in many leading evaluation frameworks. This remains true, even after what we know about the impacts of the pandemic, and despite the normative framework offered by human rights within which evaluators can assess public policy. Nonetheless, these legal standards should support the conscious reorientation of policies from needs-based to rights-based approaches, providing a lens through which policies and evaluation frameworks can be aligned with human rights.

Whence human rights in evaluation?

Zenda Ofir and Deborah Rugg (2021) argue powerfully that the pandemic presented a unique opportunity to reimagine society and its relationship to the environment, as well as to redefine our own values. One way to engage with the needed systems transformation is to begin with the fundamentals, namely, an inquiry into what human rights and HRBA entail.

Integrating human rights in the evaluation context

"Human rights" are often used to assert a multiplicity of claims. For the purposes of this discussion, however, human rights refer to the catalogue of rights recognized in international human rights law which reflect decades, sometimes centuries, of evolution and consideration by lawyers, judges, practitioners, activists, and civil society. As Hurst Hannum (2020) notes, "'[h]uman rights' may mean all things to all people, but 'international human rights law' cannot" (p. 15).

The 1948 *Universal Declaration of Human Rights* (UDHR) was the starting point for the international human rights regime in the period following the Second World War. Two decades later, the *International Covenant on Civil and Political Rights* (ICCPR) was adopted. The ICCPR is one of the core UN human rights treaties, transforming the UDHR's aspirations for civil and political rights into legal obligations. The *International Covenant on Economic, Social and Cultural Rights* (ICESCR) was adopted at the same time, and protects the rights to education, health, and adequate housing, among others. Together, the UDHR, the ICCPR and the ICESCR form the bedrock of the international human rights system and comprise the International Bill of Rights.

There are now nine core human rights instruments and more than 100 other human rights instruments. The proliferation of rights has been the result, at least in part, of the need to reformulate and innovate rights within established frameworks so that new, evolving, or more specific circumstances affecting people who had previously been left out would be addressed, with "no one left behind" in mind. These include women, children, migrants, people with disabilities, and Indigenous peoples, among others. Most of the conventions and other binding international

instruments that protect these groups, especially the more recent instruments, now integrate civil and political rights with economic, social, and cultural rights to emphasize the interdependence of human rights.

Together, the commitment to respect, protect, and fulfill human rights form complex networks of interlocking, interdependent, and indivisible human rights obligations (United Nations General Assembly, 1993). Most of the main human rights treaties, like conventions and covenants, have been ratified by a significant majority of the world's countries. And yet, few multilateral organizations have managed to integrate these areas effectively at a systems wide level.

Human rights are supposed to operate horizontally across systems, reinforcing the interdependence of rights, but they also act vertically, cascading into regional levels in the African, American, and European human rights systems, for example, and then into States via national constitutions, human rights institutions, statutes, and other legal instruments. The same rights are often reiterated at sub-national levels, in provinces, states and territories through human rights laws or codes. For example, the right to adequate housing at the international level is often translated or "domesticated" at the national level, at least in part, through human rights legislation prohibiting discrimination in housing.

There is no doubt that human rights, like all areas of law and policy, have become increasingly complex as we recognize the inherent interdependence and interconnections among different types of rights and their interaction with social and economic policies. In this respect, evaluators who are required to assess impacts on marginalized and vulnerable communities must be aware of or at least have working familiarity with international human rights instruments that affect those communities.

This very complexity gives rise to reasonable challenges to the human rights framework. The rapid evolution and proliferation of rights has resulted in what scholars like Eric Posner (2014, p. 94) in the United States and Dominique Clément (2018, p. 4) in Canada have called a "hypertrophy" or "inflation" of human rights, respectively. Critics argue that this inflation devalues laws and institutions and renders human rights less effective, diminishing the value that human rights law was designed to protect.[8] Some go further, arguing that human rights pose nothing less than a danger to democratic systems because courts become empowered to override the will of democratically elected legislatures (Greene, 2021). Even among those who are sympathetic to the human rights project, there are concerns that human rights frameworks may prove inadequate to address the central challenges to humanity in the 21st century, notably those of poverty, conflict, and the environment (Akande et al., 2020).

Compounding the problem is the reluctance in common law jurisdictions like Canada, the United States, the United Kingdom, and Australia to recognize the justiciability (i.e., the ability to go to court to enforce

claims) of economic and social rights. These rights include education, health, housing, and decent work.[9] Challenges to justiciability are largely a product of the Cold War mentality that led to the bifurcation of civil and political rights on the one hand, and economic, social, and cultural rights on the other, giving primacy to civil and political rights in the West and contributing to the view that these rights are not "real rights" but mere aspirations (Eliadis, 2018). For example, the right to adequate housing is contained in the ICESCR, but most countries treat housing as a commodity and few common law countries guarantee the right to adequate housing as a freestanding or independent right outside the area of discrimination. Shifting the discourse and the policy focus to an HRBA has started in some place, but it has been a decades-long struggle in countries like Canada. Even so, policymakers and evaluators will need to develop entirely new baseline information, targets, and indicators to align housing strategies, including market-based strategies, programs, and policies to an HRBA that aims toward the progressive realization of rights.

The "too many rights" argument has had other, insidious, impacts beyond limited justiciability and the resulting loss of access to justice. The pandemic has forced us to reconsider the way in which public policy is handled, which has implications for evaluation. For example, older people in long-term care facilities and residences experienced extremely high death rates during the pandemic. In Quebec, the mortality rate was exacerbated by a decision to move older, infected persons out of hospitals to free up beds for (younger) patients. The result was a human catastrophe as the virus rapidly spread in these facilities, whose residents were already frail and where panicked staff fled, refusing to come to work (Protecteur du citoyen, 2021). Many of these deaths were a direct result of ageism and the fact that congregate living facilities for older and infirm persons were simply not on the policy "map" as preserving hospital beds was the only consideration at play. Had an HRBA been taken, one that used an intersectional approach which considered age and disability, the death toll likely would have been much lower if patients had remained in hospitals with trained staff and appropriate protocols for infectious diseases. Such an approach would have been sensitive to needs of this very vulnerable group who had no recourse and no capacity to mobilize or advocate for itself. In short, the proponents of rights inflation arguments effectively hobble attempts to protect wider cross-sections of humanity. Ignoring HRBA also effectively undermines initiatives to ensure that more people can access justice (Des Rosiers, 2018).

Even if it is accepted that human rights should play a much stronger role in evaluation standards and ethical frameworks, and even if the *a priori* primacy of human rights may be clear in theory, it is not always obvious how to translate these lofty principles, whether current or emerging, into practice. This is especially true when setting priorities across multiple policy objectives with finite resources (Nickel, 2008). This is where HRBA

make important connections between the normative primacy of human rights and their operationalization in programs and policies.

HRBA as a solution?

HRBA originated in the UN system in the 1990s (United Nations, Sustainable Development Group, 2003) and were also used as a development tool by international NGOs (Nelson & Dorsey, 2018). HRBA forms one of the six Guiding Principles of the *United Nations Sustainable Development Cooperation Framework* (United Nations, 2019, p. 11). In its basic formulation, HRBA permits public policy design and its evaluation to integrate the foundational values of human society. As a conceptual framework, HRBA is normatively based on international human rights standards and operationally directed to promoting and protecting human rights.

HRBA is intended, first, as a fundamental and systems-organizing paradigm for programming in all sectors and in all phases of policy design and development. Second, it supports participants, citizens, civil society, and residents to contribute to their own capacity as "rights-holders" and to claim their rights against States and other duty bearers. It requires that policies and programs be aimed at reducing disparities affecting marginalized, disadvantaged, and excluded groups. Third, HRBA offers quantitative techniques such as systematic disaggregation of data by gender, disability, race, and other grounds to identify impacts at a more granular level. Failure to use such techniques progressively can perpetuate inequality and discrimination, something that has become even more obvious in the context of the pandemic (Packer & Balan, 2020).

HRBA is broadly relevant to all program and policy evaluations regardless of the sector or type of evaluation (Eliadis, 2021). The requirement that measures be *operationally* directed to promoting and protecting human rights means that policy frameworks must be reoriented, along with policy and program evaluation. This observation applied beyond the evaluation of projects that are explicitly grounded in human rights, rule of law, or other justice-oriented projects. Even relatively early on, it was clear that it was no longer "business as usual" (Raimondo et al., 2020; Sandhu et al., 2020). Some of the key elements of HRBA area that they are:

- *Equality-focused*: Outcomes should be based on substantive equality which means that policies consciously consider and prioritize groups experiencing discrimination and marginalization.
- *People-centered:* States and (sometimes) third parties owe duties to people, not programs or policies, or even to systems. Putting people first requires a shift so the rights and perspectives of those most affected or likely to need the policy or program, are taken into consideration from the outset.

- *Progressively realized:* Social, economic, and cultural rights such as education, health, transit, sanitation and clean water, and adequate social assistance depend on the principle of progressive realization, recognizing that few States can meet targets in the short term and that significant investments in infrastructure and programs are needed, to the maximum of available resources.
- *Process-oriented:* Equal importance is given to processes of policy development and meaningful participation is a key strategy to ensure the input of those most affected. Participatory approaches establish and strengthen choices and opportunities for self-development and self-fulfillment within a sustainable development framework (Eliadis, 2021).

HRBA are used across several areas of economic, social, and cultural policy to provide normative direction. For example, the use of mainstreamed gender-based analyses – a central part of the human rights framework – can be seen in "gender-based analysis plus" or GBA+ to address not only gender-related impacts but also intersections with other human rights grounds such as disability, race, 2SLGBTQI (two-spirited, lesbian, gay, bisexual, trans, queer, and intersex) status, nationality, religion, and so on. In Canada, GBA+ is used to build public policy and orient it toward women's rights while considering intersectional implications for race, disability, and other human rights grounds (Women and Gender Equality Canada, 2021). GBA+ also supports the Sustainable Development Goals (SDG) such as SDG 1 (Ending Poverty), SDG 5 (Gender Equality), SDG 8 (Economic Growth and Decent Work), and SDG 10 (Reducing Inequality), to name but a few. Most evaluation units are familiar with gender-based analyses, but the challenge now is to move that intersectional analysis to other areas.

As noted earlier, policy areas such as housing are increasingly starting to include HRBA, or variations of it, as part of regional and country-level policies.[10] In 2019, Canada enacted the *National Housing Strategy Act* (the "Act"), which established a legislative and policy framework for an HRBA to adequate housing as a fundamental right that is essential to dignity and well-being.[11] The Act aims to support improved housing outcomes and explicitly incorporates the international human rights law standards of progressive realization of the right to adequate housing, especially for those in highest housing need and those facing multiple barriers in having their housing needs met. These groups, at least in the Canadian context, include Indigenous peoples, survivors fleeing domestic violence, racialized groups, immigrants and refugees, persons experiencing homelessness, people with disabilities, those dealing with mental health and addiction issues, veterans, seniors, young adults, members of 2SLGBTQI communities, and women and gender-diverse persons within these groups.

The Act establishes a new and independent Federal Housing Advocate (the "Advocate") as the monitoring mechanism and creates a robust

monitoring framework. Given what we have learned about how the pandemic has particularly affected marginalized and vulnerable groups, evaluators will have important responsibilities to ensure HRBA going forward.

Evaluators also need to know that there are special rules that apply during pandemics or indeed during any kind of emergency. The next section discusses how human rights and HRBA become more important during times of emergency, precisely because of what we know now about the increased vulnerability of the very groups that human rights are intended to prioritize and protect.

Human rights and HRBA in times of emergency

The COVID-19 crisis presented real dilemmas: Which rights and values should be prioritized in practice? Should the right to life and health prevail over personal liberty, or is it preferable to protect freedom and economic activity? By what standards should these decisions be made? And what are the implications for evaluation? The global health emergency offered a real-world and real-time "lab" in which these choices and their consequences could be assessed in close to real time. Trying to answer the questions with standards like "efficiency," "effectiveness," or "propriety" would have offered little in the way of assistance in making those choices.

Human rights law and HRBA, on the other hand, do have something to say about such choices. The UN Office of the High Commission for Human Rights (OHCHR, n.d.) has taken the position that addressing the pandemic must include "not only the medical aspects of pandemics, but also the human rights and gender-specific impacts of measures, especially for vulnerable and marginalised communities." That is because, during emergencies, rights are more at risk than ever.

Protecting human life and human security should be a paramount consideration, as the right to life (unlike most other civil and political rights or social and cultural rights) cannot be derogated from, even in times of emergency. Life is an obvious precondition to all other rights, including the capacity to participate in economic activity. This ex-ante legal reasoning is supported by empirical data: those countries which actively took measures, sometimes extreme measures, to suppress disease were more likely to succeed at both preserving life and supporting the economy. According to an analysis conducted by Alvelda et al. (2020), imposing restrictions like lockdowns and other measures to suppress disease may have been unpopular with some policymakers because of the impacts on the economy but such measures proved to be more effective in both suppressing the disease and in protecting the economy than measures that simply protected the economy.

Limiting economic damage caused by the pandemic starts and ends with controlling the spread of the virus. Data from dozens of countries across the world suggests that no country can prevent the economic damage

without first addressing the pandemic that causes it: "those countries, like the United States, that invested in economic stimulus while allowing the virus to continue proliferating continue to suffer from unabated community transmission, deepening the economic damage quarter after quarter as the virus spreads" (Alvelda et al., 2020).

First, and as previously mentioned, vulnerable and marginalized groups were disproportionately impacted and even endangered by blanket measures that were intended to protect the general population. Second, vigilance was required to ensure access to information and participation in democratic processes, especially so that people and groups whose voices are often drowned out in times of emergency can be heard (Amnesty International Canada, 2020). Such measures should also be evaluated and monitored during emergencies. A strong participatory element at the design and monitoring stages can ensure that affected groups are considered early in the process and that evaluation of public policies considers not only the statistics that affect the general population, but also disaggregated information to identify and address disproportionate impacts on equality-seeking groups.

Being able to evaluate measures that are invoked during times of emergency requires an understanding of how human rights operate in times of emergency, and this has real implications for policy and program design and evaluation.

Public measures derogating from human rights

This section provides an overview of the rules that states must follow during public emergencies. Evaluators are not usually lawyers, but they should be aware of basic conditions of lawfulness and respect for the rule of law. According to Eric Richardson and Colleen Devine (2020), the legal consequences of failing to respect human rights standards in times of emergency create a set of specific harms. Evaluators must be sensitive to such risks, if only because of the evaluator's ethical imperative to "do no harm" (Richardson & Devine, 2020).

Under the ICCPR, derogations from civil and political rights in times of emergency must only be in response to "threats to the life of the nation." The Siracusa Principles, which provide interpretive guidance to the ICCPR, provide that the such a threat must (1) be "actual or imminent"; (2) "affect the whole of the population and either the whole or part of the territory of the State"; and (3) "threaten the physical integrity of the population, the political independence or the territorial integrity of the State or the existence or basic functioning of institutions indispensable to ensure and protect the rights recognized in the Covenant" (United Nations, 1984). Shortly after the pandemic began, the United Nations Human Rights' Committee recognized that a "public emergency" includes public health emergencies: "States parties confronting the threat of widespread

contagion may resort, on a temporary basis, to exceptional emergency powers and invoke their right of derogation from the Covenant" (United Nations, Human Rights Committee, 2020). However, such measures must be strictly tailored to the exigencies of the situation.

Derogations from human rights are restricted during emergencies to ensure that people are protected from abuses of government power during periods of reduced democratic oversight. Transparency and accountability are eroded during public emergencies: legislatures are suspended, judicial systems are less effective and accessible, and courts tend to show deference to the power of the executive. At the time of writing, for example, few court cases outside the United States had successfully struck down the major planks of government measures against the pandemic. One of the exceptions is Spain, where there was a successful challenge against most public health measures before Spain's Constitutional Court in July 2021. However, the crux of the decision was that the government had chosen the wrong legal mechanism to derogate from fundamental rights, and not that the emergency measures were themselves unjustified (Spain, 2021).

The very measures adopted to limit infections and illness and protect rights – including lockdowns, travel restrictions, quarantines, masking, and curfews – can also limit human rights. To avoid this result, Article 4 of the ICESCR, which protects economic, social, and cultural rights (including the right to the "highest standard of physical and mental health"), provides that states may subject such rights to only such limitations as are "determined by law, only insofar as this may be compatible with the general nature of these rights and solely for the purpose of promoting the general welfare in a democratic society."

Article 4(3) of the ICCPR requires countries which plan to derogate from rights during times of emergency to immediately "inform the other parties to [the ICCPR], through the intermediary of the Secretary-General of the United Nations, of the provisions from which it is derogated and of the reasons by which it was actuated." The obligation to provide notice is not intended to be a mere formality (McGoldrick, 2014, p. 422). Nonetheless, as of December 31, 2021, only 24 of the United Nations' 193 Member States countries had declared a state of emergency related to the COVID-19 pandemic and communicated it to the Secretary-General as required by Article 4(3) of the ICCPR (see also United Nations, Human Rights Committee, 2020). Failing to notify the international community of existing derogations restricts the capacity of the international community to engage in oversight of human rights violations (Richardson & Devine, 2020, p. 124).

There are certain rights that carry with them the possibility of restrictions or limitations, without the need to declare a public emergency or to provide notification. Mobility rights, for example, can be limited for reasons of public health, a restriction that is built into the ICCPR. The same applies to the rights to freedom of religion, expression, association, and

peaceful assembly. Again, evaluators need to be aware of these nuances to properly assess the lawfulness of emergency public measures.[12] The principle of *legality* requires that the measures taken in response to an emergency must be "prescribed by law." That means that there needs to be a validly passed or enacted legal instrument, such as a statute or regulation. In cases of emergency where the legislature may have been suspended, governments use orders-in-council or other executive orders.[13] These are fine, provided they have a valid legal basis, such as an authorizing statute. Administrative and policy documents, on the other hand, are not likely to be compatible with the principle of legality. Executive orders, provided they are issued under statutory authority as mentioned above, will comply (Richardson & Devine, 2020, pp. 114–115).

Virtually every nation state has the power to invoke emergency or public health laws that allow for such special measures. Laws may operate at the national and subnational levels, often for jurisdictional reasons. In Canada, for instance, the federal government chose not to invoke federal emergency legislation for reasons related to public health specifically (although it did invoke the *Emergencies Act* in February 2022 after right-wing protests occupied parts of Ottawa, Canada's capital, as well as areas such as bridges which are considered critical infrastructure). However, it did invoke the *Quarantine Act*, among other laws that aimed mainly at controlling entry into the country. The delivery of health care in Canada is mainly a provincial responsibility, so each province used their own powers deriving from emergency and/or public health legislation.[14]

The point here is that national constitutions usually contain some sort of power to balance competing rights with each other, or even to weigh rights against other values and norms, including in times of emergency. Section 1 of the *Canadian Charter of Rights and Freedoms*, for example, allows for reasonable limits on rights that are demonstrably justifiable in a free and democratic society. In Europe, States are given a margin of appreciation to decide on legal measures that may affect or restrict rights.

Several of the criteria related to the validity of public measures in response to threats are squarely within the competence of evaluators because most rely on empirical data, rights assessment, and risk management. They include an assessment of whether the measures are:

- *Necessary* to address the public health emergency;
- *proportionate responses* to the emergency in the sense that they are limited in scope (e.g., duration, geographic territory, and in terms of substantive rights limitations);
- *compliant with the principle of non-discrimination* and of the rule regarding the non-derogation of rights, notably the right to life; and
- designed to be *temporary* in nature, with a view to an eventual return to normalcy and the restoration of rights and freedoms (ICCPR).

Necessity and proportionality are especially well-suited to evaluators' expertise, since they require an assessment of evidence about, for example, whether epidemiological data or at least an application of the precautionary principle may provide support for certain public measures, as well as their effectiveness, which in turn justifies their ongoing or mitigated use.

In summary, in times of emergency, governments often invoke measures whose effectiveness must be assessed not only in terms of infection rates, or economic impacts, but also in relation to compliance with human rights norms and the special rules that apply when governments seek to derogate from fundamental human rights. One of the most essential of these rules is that even in times of emergency, public health and social measures cannot under any circumstances operate in a discriminatory manner.

Principle of non-discrimination

Article 2 of the ICCPR provides that rights must be protected without distinction of any kind, such as "race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status." Even in public emergencies, discrimination is never permitted when based solely on race, color, sex, language, religion, or social origin (Article 4(1)).

We now know that the burden of the public health crisis has fallen on marginalized people, often because of their pre-existing vulnerabilities, resulting in disproportionate impacts of public and health measures. The example noted earlier with respect to the transfer of older patients to long-term care facilities was based *solely* on age (younger people were not transferred). In the example cited earlier with respect to people living in long-term care homes, seniors were subjected to measures designed to support the public health system by freeing up hospital beds in primary care hospitals; underpinning those measures was the ageist and ableist assumption that such people did not "need" acute care. When the authorities transferred many older people from the hospitals to long-term care facilities, the latter became death traps (Protecteur du citoyen, 2021).

In other cases, failing to center human rights within public policy has permitted responses that have worsened these vulnerabilities even if the discrimination may not have been solely based on the enumerable factors. The principle of non-discrimination thus assumes critical significance for marginalized groups whose circumstances have been worsened by the pandemic and by the measures that have been constructed to combat it. A study by Camelia Gianella et al. (2020) on human rights during Peru's response to the pandemic, noted that "the Covid 19 pandemic has brought attention to deep inequalities within the system, including gender and ethnic inequalities" (p. 318). In the United States, the data has shown gross disparities in mortality rates during the pandemic for racialized groups, especially Black and Indigenous peoples (APM Research Lab, 2021). In the United States, the Centers for Disease Control and Prevention have

reported that Indigenous and Black individuals faced age-adjusted rates of 3.5 and 2.8 times the risk of being hospitalized for the COVID-19 infection, respectively (Centers for Disease Control and Prevention, 2021). In short, there are myriad ways in which systemic and structural forms of discrimination and racism have affected minorities groups, including racialized people who were exposed to the virus because of the jobs they held or the neighborhoods in which they live. Women were also disproportionately affected, especially racialized women (Mattioli et al., 2021). These outcomes have been associated with factors such as stress and additional caregiving responsibilities which have increased susceptibility to infection, aggravated by existing co-morbidities. There is growing evidence at the global level that domestic violence has also increased (Peterman & O'Donnell, 2020; Usher et al., 2020).

People who experience homelessness also experience severe and adverse impacts. Many were unable to move to safe locations or access a home during quarantine or curfew. In Canada, a legal clinic had to go to court to seek an exemption from a night-time curfew for people living in the street after the government of Quebec refused to make exceptions. A man who was living on the street and who sought to avoid the police had sheltered in a portable toilet and froze to death during the harsh Montreal winter in early 2021, triggering a public outcry. The Superior Court of Quebec granted the exemption in the case *Clinique juridique itinérante c. Procureur général du Québec* (Quebec Superior Court, 2021). The Quebec decision shows how some emergency measures developed for the general welfare of citizens can be inherently insensitive to the conditions of vulnerable groups and poor communities.

These are but a few examples that illustrate why understanding discrimination is vital for policymakers and for evaluators alike, regardless of whether a particular evaluation mandate explicitly refers to "human rights" in its title or is part of a social justice or rule of law project.

Conclusion: Substantive transformations needed

Human rights are, or should be, especially relevant to evaluating public policy, particularly in times of emergency. This chapter has argued that human rights should provide the normative ballast for public health practice during turbulent times. It is important to observe criteria such as "do no harm," "do good," and "propriety." But it is essential to ensure that programs and policies are always lawful and consistent with HRBA.

Evaluators should assess the extent to which interventions comply with existing human rights standards, to be aware of emerging rights and norms, and show careful attention to the needs of those who are already marginalized and vulnerable. Evaluation becomes a tool to connect human rights standards into policy design and evaluation. It emphasizes

"whole systems" thinking, reflecting the legal principle that human rights are interdependent and operate in a non-hierarchical manner.

Human rights and HRBA not only prioritize human life but also provide rules to manage rights, conflicts, and values trade-offs. Additional oversight and engagement measures can also help to support participation and democratic deliberation during times of emergency and serve to minimize the risks of normalization beyond the pandemic.

That is why human rights should be strengthened, not diminished, during times of emergency. Failure to assess lawfulness may also increase the likelihood that emergency measures will continue without scrutiny. Finally, the determination of whether the measures are necessary, proportionate, and adhere to the principles of non-discrimination is especially well-suited to evaluation practice. If evaluators are to maintain their own value proposition, the systems transformation that are needed will necessarily place human rights at the top of the pyramid of values and standards within their professional practice.

Notes

1 The chapter was prepared with the assistance of Marial Aramburu, who was an MPP candidate at the Max Bell School of Public Policy, McGill University at the time of writing and is currently with the United Nations World Food Programme. Many thanks to Indran Naidoo and Ray Rist for their thoughtful and supportive feedback on this chapter and to Fionnuala Tennyson and Shaun Ryan for their review and editing of the text. Errors are my own.

2 Well-known examples include female genital cutting, child marriage and girl-kidnapping, impunity for sexual assault, and marital rape. In addition, discriminatory property, contract, family and succession laws severely limit women's rights in virtually all spheres of life in a wide range of countries.

3 The analysis was conducted by Mariel Aramburu who provided research support and assistance for this book. See note 1, above.

4 African Development Bank, Independent Development Evaluation; Asian Development Bank, Independent Evaluation; Inter-American Development Bank, Evaluation and Oversight and the World Bank, Independent Evaluation Group.

5 UN Development Programme (UNDP), Independent Evaluation Office; UN Evaluation Group (UNEG), UN Women, IFAD Independent Office of Evaluation.

6 African Evaluation Association (AfrEA); Asia-Pacific Evaluation Association; Caribbean Evaluators International (CaribEval); Community of Evaluators (CoE) in South Asia; Eurasian Alliance of National Evaluation Associations (EvalEurasia); European Evaluation Society; Evaluators Network of the Middle East and North Africa (EvalMENA); Red de Seguimiento, Evaluacion y Sistematización en America Latina y el Caribe (ReLAC 2.0).

7 Australian Evaluation Society; American Evaluation Association; Canadian Evaluation Society; Ghana Monitoring and Evaluation Forum; Lebanese Evaluation Association (LebEval); Academia Nacional de Evaluadores de Mexico (ACEVAL); South African Monitoring and Evaluation Association (SAMEA); Réseau Tunisien d'Evaluation (RTE).

8 There are several authors who have developed criticisms of "rights talk," the rise of human rights and the inherent limits on human rights (see also Henkin, 1990; Ignatieff, 2001).

9 In the Canadian context, for example, see *Gosselin v. Québec (Attorney General)* (2002), SCC 84, where the court refused to extend the rights to equality or to life, liberty and security of the person to an older woman on social assistance, and *Tanudjaja, et al. v. Attorney General of Canada, et al.* (2014), 326 O.A.C. 257 (application for leave by the SCC dismissed June 25, 2015), which denied the right to housing under the Canadian Charter of Rights and Freedoms.

10 In Latin America, for example, see Yamin and Frisancho (2014). At the international level, there is a developing normative approach to the right to adequate housing, including at the UN (2020).

11 The 2017 federal National Housing Strategy culminated in the *National Housing Strategy Act*, S.C. (2019), c 29, which created a legislative right to adequate housing for the first time in in Canada's history.

12 There is the additional possibility that a public measure may be otherwise valid by reason of a reservation, understanding, or declaration made by States in accordance with international law. These mechanisms are beyond the purview of this paper, but evaluators should be aware of this possibility.

13 The term "prescribed by law" is contained in the ICCPR, and the requirement of a valid legal instrument as a condition of legality is supported by General Comment 37 and by the Siracusa Principles. Human Rights Committee ["HRC"], General Comment 37 art. 21, para 39, U.N. Doc. CCPR/C/37 (July 23, 2020) [General Comment No. 37].

14 In Quebec, for example, which experienced the largest number of cases in the country and the first wave of the pandemic, the government relied principally on *the Public Health Act*, RSQ c S-2.2. and to a much lesser extent on the *Civil Protection Act c S-23*.

References

Legal sources

Charter of the United Nations, 26 June, Can TS 1945 No 7. (1945). https://www.un.org/en/about-us/un-charter/full-text

Clinique juridique itinérante c. Procureur général du Québec, Q.C.C.S. 182 (CanLII) (Can.). (2021). https://www.canlii.org/fr/qc/qccs/doc/2021/2021qccs182/2021qccs182.html?autocompleteStr=Clinique%20juridique%20itin%C3%A9rante%20c.%20Procureur%20g%C3%A9n%C3%A9ral%20du%20Qu%C3%A9bec%2C%202021%20Q.C.C.S.%20182%20(CanLII)%20&autocompletePos=1

Gosselin v. Québec (Attorney General), S.C.C. 84. (2002). https://scc-csc.lexum.com/scc-csc/scc-csc/en/item/2027/index.do

International Covenant on Civil and Political Rights. (Dec. 16). 999 UNTS 171. (1966) https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights.

International Covenant on Economic, Social and Cultural Rights. UNTS 993. (1966) https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-economic-social-and-cultural-rights. https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-economic-social-and-cultural-rights

National Housing Strategy Act, S.C., c 29. (2019). https://laws-lois.justice.gc.ca/eng/acts/n-11.2/FullText.html

Public Health Act, R.S.Q. c S-2.2. https://www.legisquebec.gouv.qc.ca/en/document/cs/s-2.2

Tanudjaja, et al. v. Attorney General of Canada, et al., 326 O.A.C. 257. (2014). https://scc-csc.lexum.com/scc-csc/scc-l-csc-a/en/item/15426/index.do

Tribunal Constitucional [Constitutional Court], Madrid. *Appeal of Unconstitutionality*, Agencia Estatal Boletín del Estado, Sentencia 148/2021. 14 July. (Spain). (2021). http://hj.tribunalconstitucional.es/HJ/en/Resolucion/Show/26778

United Nations General Assembly. Vienna Declaration and Programme of Action, 12 July 1993, A/CONF.157/23. (1993). https://www.ohchr.org/en/instruments-mechanisms/instruments/vienna-declaration-and-programme-action

Secondary sources

Akande, D., Kuosmanen, J., McDermott, H., & Roser, D. (Eds.) (2020). *Human rights and 21st century challenges: Poverty, conflict, and the environment*. Oxford University Press.

Alvelda, P., Ferguson, T., & Mallery, J. C. (2020, November 18). *To save the economy, save people first*. Institute for New Economic Thinking. https://www.ineteconomics.org/perspectives/blog/to-save-the-economy-save-people-first

American Evaluation Association. (2018). *Guiding principles for evaluators*. https://www.eval.org/About/Guiding-Principles

Amnesty International Canada. (2020). *A call for human rights oversight of government responses to the COVID-19 pandemic*. https://amnesty.ca/wp-content/uploads/2021/08/COVID-and-human-rights-oversight-public-statement-FINAL_0.pdf

APM Research Lab. (2021, March 5). *The color of coronavirus: COVID-19 deaths by race and ethnicity in the U.S.* https://www.apmresearchlab.org/covid/deaths-by-race

Centers for Disease Control and Prevention. (2021). *Hospitalization rates per 100,000 population by age and race and ethnicity — COVID-NET, March 1, 2020–July 17, 2021*. https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/racial-ethnic-disparities/disparities-hospitalization.html

Clément, D. (2018). *Debating rights inflation in Canada: A sociology of human rights*. Wilfrid Laurier Press.

Connor, J., Madhavan, S., Mokashi, M., Amanuel, H., Johnson, N. R., Pace, L. E., & Bartz, D. (2020). Health risks and outcomes that disproportionately affect women during the Covid-19 pandemic: A review. *Social Science & Medicine*, *266*, 113364. https://doi.org/10.1016/j.socscimed.2020.113364

Cutler, D. M., & Summers, L. H. (2020). The COVID-19 pandemic and the $16 trillion virus. *JAMA Network*, *324*(15), 1495–1496. https://doi.org/10.1001/jama.2020.19759

Des Rosiers, N. (2018). The right investment in rights. In D. Clément (Ed.), *Debating rights inflation in Canada: A sociology of human rights* (pp. 79–96). Wilfrid Laurier Press.

Eliadis, P. (2018). Too many rights? In D. Clément (Ed.), *Debating rights inflation in Canada: A sociology of human rights* (pp. 97–120). Wilfrid Laurier Press.

Eliadis, P. (2021). Mind the gaps: Integrating human rights and sustainable development goals into evaluation practice. In A. Paulson & M. Palenberg (Eds.). *The realpolitik of evaluation: Why supply and demand rarely intersect* (pp. 123–325). Routledge. https://doi-org.proxy3.library.mcgill.ca/10.4324/9781003005162

Forss, K., & Marra, M. (Eds.). (2014). *Speaking justice to power: Ethical and methodological challenges for evaluators*. Routledge. https://doi.org/10.4324/9781315130132

Gianella, C., Iguiñiz-Romero, R., Romero, M. F., & Gideon, J. (2020). Good health indicators are not enough: Lessons from COVID-19 in Peru. *Health and Human Rights Journal*, *22*(2), 317–320. https://www.hhrjournal.org/2020/06/good-health-indicators-are-not-enough-lessons-from-covid-19-in-peru/

Greene, J. (2021). *How human rights went wrong: Why our obsession with rights is tearing America apart*. Houghton Mifflin Harcourt.

Guterres, A. (2020). *The highest aspiration: A call to action for human rights*. United Nations. https://www.un.org/sg/sites/www.un.org.sg/files/atoms/files/The_Highest_Asperation_A_Call_To_Action_For_Human_Right_English.pdf

Hannum, H. (2020). Reinvigorating human rights for the twenty-first century. In D. Akande, J. Kuosmanen, H. McDermott, & D. Roser (Eds.). *Human rights and 21st century challenges: Poverty, conflict, and the environment* (pp. 13–138). Oxford University Press.

Henkin, L. (1990). *The age of rights*. Columbia University Press.

Ignatieff, M. (2001). *Human rights as politics and idolatry* (A. Gutman, Ed.). Princeton University Press.

Kjaerum, M., Davis, M. F., & Lyons, A. (Eds.). (2021). *COVID-19 and human rights*. Routledge. https://doi.org/10.4324/9781003139140

Landry, M. D., Van den Bergh, G., Hjelle, K. M., Jalovcic, D., & Tuntland, H. K. (2020). Betrayal of trust? The impact of the COVID-19 global pandemic on older persons. *Journal of Applied Gerontology*, *39*(7), 687–689. https://doi.org/10.1177/0733464820924131

Mattioli, A. V., Sciomer, S., Maffei, S., & Gallina, S. (2021). Lifestyle and stress management in women during COVID-19 pandemic: Impact on cardiovascular risk burden. *American Journal of Lifestyle Medicine*, *15*(3), 356–359. https://doi.org/10.1177/1559827620981014

McGoldrick, D. (2014). The interface between public emergency powers and international law. *International Journal of Constitutional Law*, *2*(2), 380. https://doi.org/10.1093/icon/2.2.380

Negrini, S., Grabljevec, K., Boldrini, P., Kiekens, C., Moslavac, S., Zampolini, M., & Christodoulou, N. (2020). Up to 2.2 million people experiencing disability suffer collateral damage each day of COVID-19 lockdown in Europe. *European Journal of Physical and Rehabilitation Medicine*, *56*(3), 361–365. https://doi.org/10.23736/S1973-9087.20.06361-3

Nelson, P. J., & Dorsey, E. (2018). Who practices rights-based development? A progress report on work at the nexus of human rights and development. *World Development*, *104*, 97–107. https://www.sciencedirect.com/science/article/pii/S0305750X17303686

Nickel, J. (2008). Rethinking indivisibility: Towards a theory of supporting relations between human rights. *Human Rights Quarterly*, *30*(4), 984–1001. https://doi.org/10.1353/hrq.0.0046

OECD DAC Network on Development Evaluation. (2010). Evaluation Development Co-Operation: Summary of Key Norms and Standards (2nd ed.) https://www.oecd.org/development/evaluation/dcdndep/41612905.pdf

Office of the High Commissioner for Human Rights. (n.d.). *COVID-19 guidance: Access to health care*. https://www.ohchr.org/EN/NewsEvents/Pages/COVID19Guidance.aspx

Office of Internal Oversight Services. (2020). *Synthesis of guidelines for UN evaluation under COVID-19*. United Nations Evaluation Group. http://www.uneval.org/document/detail/2863

Ofir, Z., & Rugg, D. (2021). Transforming evaluation for times of global transformation. *American Journal of Evaluation*, *42*(1), 47–52. https://doi.org/10.1177%2F1098214020979070

Organisation for Economic Co-operation and Development – Development Assistance Committee Network on Development Evaluation. (2010). *Evaluating*

development co-operating summary of key norms and standards (2nd ed.). http://www.oecd.org/development/evaluation/dcdndep/41612905.pdf

Packer, J., & Balan, S. (2020, July 27). *A genuine human rights-based approach for our post-pandemic future*. OpenGlobalRights. https://www.openglobalrights.org/genuine-human-rights-based-approach-for-post-pandemic-future/

Patton, M. Q. (2020). Evaluation criteria for evaluating transformation: Implications for the coronavirus pandemic and the global climate emergency. *American Journal of Evaluation*, *42*(1), 53–89. https://doi.org/10.1177/1098214020933689

Peterman, A., & O'Donnell, M. (2020, September 23). *COVID-19 and violence against women and children: A second research roundup*. Centre for Global Development. https://www.cgdev.org/publication/covid-19-and-violence-against-women-and-children-second-research-round

Posner, E. (2014). *The twilight of international human rights law*. Oxford University Press.

Prevent Epidemics. (2020, May 12). *Assessing COVID-19 resurgence*. https://preventepidemics.org/covid19/science/insights/assessing-covid-19-resurgence/

Protecteur du Citoyen. (2021) *La COVID-19 dans les CHSLD durant la première vague de la pandémie Cibler les causes de la crise, agir, se souvenir*. Québec Ombudsperson. https://protecteurducitoyen.qc.ca/sites/default/files/2021-11/rapport-special-chsld-premiere-vague-covid-19.pdf

Raimondo, E., Vaessen, J., & Branco, M. (2020, April 22). *Adapting evaluation designs in times of COVID-19 (coronavirus): Four questions to guide decisions*. World Bank Independent Evaluation Group. https://ieg.worldbankgroup.org/blog/adapting-evaluation-designs-times-covid-19-coronavirus-four-questions-guide-decisions

Richardson, E., & Devine, C. (2020). Emergencies end eventually: How to better analyze human Rights restrictions sparked by the COVID-19 pandemic under the International Covenant on Civil and Political Rights. *Michigan Journal of International Law*, *42*(1), 105–176. https://doi.org/10.36642/mjil.42.1

Sandhu, S. C., Prakash, B., & Chaneco, M. A. (2020, August 10). *Learning from the past for a bright post-pandemic future*. Asian Development Blog. https://blogs.adb.org/blog/learning-from-past-for-bright-post-pandemic-future

Stame, N. (2018). Strengthening the ethical expertise of evaluators. *Evaluation*, *24*(4), 438–451. https://doi.org/10.1177%2F1356389018804942

United Nations, Economic and Social Council. (1984). *Siracusa principles on the limitation and derogation provisions in the International Covenant on Civil and Political Rights*. In E/CN.4/1985/4. United Nations Commission on Human Rights. https://www.refworld.org/docid/4672bc122.html

United Nations, Human Rights Committee. (2020). *Statement on derogations from the Covenant*. United Nations Digital Library. https://digitallibrary.un.org/record/3863948

United Nations, Human Rights Council. (2007). *Report of the Special Rapporteur on adequate housing as a component of the right to an adequate standard of living, Miloon Kothari*. In A/HRC/4/18 at Annex II. https://documents-dds-ny.un.org/doc/UNDOC/GEN/G07/106/28/PDF/G0710628.pdf?OpenElement

United Nations, Human Rights Council. (2019). *Guidelines for the implementation of the right to adequate housing*. In A/HRC/43/43. United Nations General Assembly. https://documents-dds-ny.un.org/doc/UNDOC/GEN/G19/353/90/PDF/G1935390.pdf?OpenElement

United Nations, Sustainable Development Group. (2003, September). *The human rights based approach to development cooperation towards a common understanding among UN agencies*. https://undg.org/document/the-human-rights-based-approach-to-development-cooperation-towards-a-common-understanding-among-un-agencies

United Nations, Women. (2022). *Pocket tool for managing evaluation during the COVID-19 pandemic*. https://www.unwomen.org/en/digital-library/publications/2020/05/pocket-tool-for-managing-evaluation-during-the-covid-19-pandemic

Usher, K., Bhullar, N., Durkin, J., Gyamfi, N., & Jackson, D. (2020). Family violence and COVID-19: Increased vulnerability and reduced options for support. *International Journal of Mental Health Nursing*, *29*(4), 549–552. https://doi.org/10.1111/inm.12735

van den Berg, R. D. (2022). Ethical guidance: Enabling evaluation's role in tackling bad and doing good. In R. D. van den Berg, P. Hawkins, & N. Stame (Eds.), *Ethics for evaluation beyond "doing no harm" to "tackling bad" and "doing good"* (pp. 17–44). Routledge.

van den Berg, R. D., Hawkins, P., & Stame, N. (Eds.). (2022). *Ethics for evaluation beyond "doing no harm" to "tackling bad" and "doing good"*. Routledge.

Women and Gender Equality Canada. (2021). *Gender-based analysis plus (GBA+)*. Government of Canada. https://women-gender-equality.canada.ca/en/gender-based-analysis-plus.html

Yamin, A. E., & Frisancho, A. (2014). Human-rights-based approaches to health in Latin America. *The Lancet*, *395*, 9975. https://doi.org/10.1016/S0140-6736(14)61280-0

Yarbrough, D. B., Shuhla, L. M., Hopson, R. K., & Caruthers, F. A. (2010). *Program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Sage Publishers.

Yeyati, E. L., & Filippini, F. (2021, 12 May). *Pandemic divergence: The social and economic costs of Covid-19*. Centre for Economic Policy Research. https://voxeu.org/article/social-and-economic-costs-covid-19

4 Knowledge Production in a Pandemic

Supporting Accountability at Pace in the United Kingdom and Canada

Jeremy Lonsdale and Maria Barrados

Knowledge is an essential prerequisite for accountability. It is the basis on which an account can be given, and that account scrutinized. Accountability knowledge can be created and analysed *within* governments by evaluators and internal auditors, *for* governments by evaluators and auditors (both internal and external), and for legislatures, by external financial and performance auditors whose studies have many features of evaluations. As demonstrated in detail in the recent book *Crossover of Audit and Evaluation Practices*, whilst the approaches and perspectives of audit and evaluation are distinct, there are marked common features and significant crossovers in practice, providing opportunities for cross-disciplinary learning and exchange (Barrados & Lonsdale, 2020).

This chapter examines how two state audit institutions (SAIs) – the National Audit Office (NAO) in the United Kingdom (UK) and the Office of the Auditor General (OAG) in Canada – have sought to produce useful knowledge for the purposes of accountability and learning in the exceptionally testing circumstances of the COVID-19 pandemic during 2020 and 2021. In particular, it considers how the more evaluative and reflective work of SAIs – performance audits, investigations, and "lessons learned" outputs – has been used to provide valuable information and knowledge to government officials and legislators at pace, in an environment in which *doing* (such as keeping people safe, securing and deploying protective equipment, and setting up unemployment relief measures), rather than *recording and rendering an account to others*, has been the highest priority for governments and those on the "front lines" of the pandemic. Such work nevertheless serves an important purpose if societies are to learn from the experience of the pandemic. If newspapers are sometimes considered to be the "first draft of history," it is possible that some of these audit reports are the preliminary scoping papers and useful evidence sources for the inevitable inquiries into the handling of the pandemic which lie ahead.

To De Bruijn (2007, p. 4), accountability "is a form of communication and requires the information that professional organizations have available to be reduced and aggregated." This reduction and aggregation can be a

DOI: 10.4324/9781003376316-5

threat to the understanding of performance, over-simplifying complex activities, and presenting an unfair or partial impression of what has been achieved and in which circumstances. As a result, in discussing the crucial characteristics of information for accountability, John Mayne (2007) highlighted it should be *credible*, *relevant*, and *timely*. Lonsdale and Mayne (2005) also identified the particular importance of *accuracy* in one key form of knowledge – performance audit reports – used for accountability purposes. In addition, in recent years, much effort has gone into improving the *presentation* of information for public reporting in an attempt to enhance understanding. Preparation of information provided to promote accountability has traditionally required time, the careful exposure of draft reports, informed debate involving the managers whose work is being assessed, and a dynamic process of learning through trial and error (De Bruijn, 2007).

All these characteristics have been threatened by the experience of the COVID-19 pandemic, which affected the entire world from early 2020. The collection and assurance of performance data, the ability to undertake site inspections or audits for verification purposes, to conduct interviews and group discussions to gather contextual evidence and insights, as well as the availability of officials to explain their actions, which are all basic aspects of audit, evaluation, and accountability processes, have been severely hampered, delayed, postponed, or abandoned. Such challenges raise interesting questions about how evaluation, and in the case of this chapter, performance audit undertaken by independent SAIs for many legislatures, has been required to adapt (Lonsdale et al., 2011). The two case examples illustrate how two of these offices not only have used the flexibilities that come with their powers but also outline the challenges they have faced.

At the time of writing, the COVID-19 pandemic is far from over, but it is clear that it has accentuated, deepened, or accelerated numerous existing aspects of the ways in which our societies operate. This has been evident, for example, in concerns about inequality (as the most vulnerable have been affected disproportionately); how development and approval processes for vaccines were carried out far faster than normal; and how the use of certain technology (e.g., the Zoom platform) which had relatively low levels of take-up before 2020 was suddenly utilized to allow personal and business contact during lockdown. Concerns have also been raised about how public funds have been spent on a scale previously considered unimaginable for, amongst other things, furlough schemes, supporting whole sectors of economies, and the procurement and production of unprecedented quantities of personal protective equipment, ventilators, and vaccines. Essential activities that normally proceeded slowly were completed quickly without much of the surrounding processes previously considered essential and – to the alarm of some – normal legislative scrutiny mechanisms were also suspended.

Arguably, the use of shortcuts and flexibilities around rules on the use of public funds are all challenges to the routine, rules-based environment of which SAIs are a significant part and have only increased the need for clarity about what has happened and hence the value of reliable and relevant knowledge. At the same time, the speed and scale of emergency action in Canada and the UK have led quickly to concern at the scale of public expenditure growth and fear of fraud, for example, in employment programmes. In the UK, it has also resulted in accusations of procurement irregularities (including suggestions of conflicts of interest and an absence of transparency around contracting) and considerable uncertainty about value for money on test and trace systems, which have been of limited utility (Comptroller and Auditor General, 2020b; 2021a).

In these circumstances, parliamentarians, and the media – twin forces for accountability, and both users of performance information generated by performance auditors and evaluators have – struggled to keep up with what has been happening in a fast-moving and, at times, highly politicized environment, in order to play their roles in scrutinizing the performance of government. There has been pressure in the UK for a formal public inquiry into the decisions that have been made which, at the time of writing, had been resisted by government as premature. In Ontario, an independent commission has already been established and made its report into COVID-19 and long-term care, given that the greatest proportions of deaths were residents in these settings (Marrocco et al., 2021). In the absence of a comprehensive examination, demands for accurate, independent, and trusted data on specific aspects of government spending have been very strong, and there has been much interest in the outputs of performance auditors, whose ability to evaluate and comment on public programmes remain in place.

The importance of environmental factors

In the 2020 book *Crossover of Audit and Evaluation Practices*, Lonsdale examined the conduct of auditing in changing times, highlighting how environmental factors played a significant role in shaping both *how* auditors did their work and *what* work they did. Focusing on the UK NAO, Lonsdale suggested that the external environment had required the NAO to work more quickly and adapt; had led to greater emphasis on assisting government, as well as fulfilling its traditional role of supporting Parliament; had increased the importance of being topical, relevant, and timely; had encouraged a greater cross-government perspective; and had forced it to adapt its ways of working. Turning to the impact of the environment on *what* work the NAO did, Lonsdale identified several trends. These included greater diversity in its outputs; a wider focus of performance audits in terms of the organizations and subjects covered; and a greater examination of more contentious and sensitive material, including commercially and politically sensitive subjects (Lonsdale, 2020).

In conclusion, Lonsdale (2020) commented that "the NAO cannot, and clearly does not, ignore the contemporary environment or environments in which it operates" (p. 128). Thus, it is not surprising that the unprecedented events of 2020 and 2021 created a dramatically different environment and have had an even more profound effect on audit work than the relatively slower-paced evolutionary changes to the type of reporting witnessed in the past. The next section of this chapter examines the impact of COVID-19 on performance audit work.

The impact of COVID-19 on performance audit work

United Kingdom

The NAO is independent of both the Parliament and the government. It undertakes the financial and performance audit of central government bodies, as well as overseeing the audit of local government. It reports to Parliament, and its outputs – particularly, performance audit (or, to the use the UK terminology, "value for money") reports – are used as the basis of the hearings of the cross-party Public Accounts Committee (PAC) within the House of Commons. The NAO has statutory powers of access to the information that it needs for its work and has freedom to report.

Like all organizations, the NAO was forced to adapt its working practices immediately and even prior to the official lockdown on March 23, 2020, staff were told to work at home. This continued for some 18 months, with limited exceptions for office working due to health reasons and unavoidable audit visits such as stock-checking purposes. Subsequently, more staff returned to office working, but many moved to hybrid office-home working. Financial statement audits were completed in 2020 and 2021, albeit in some cases to delayed timetables, but also with some efficiency gains where, for example, online meetings replaced physical meetings which previously required often extensive travel.

The results of these audits underlined the impact faced right across government; for example, the audit opinions in many cases highlighted uncertainty in income streams or the unplanned expenditure (given the accounting year-end in the UK is March 31, the impact was felt even more in the 2020–2021 audits). Performance audits continued largely unaffected, with teams and those subject to audit adapting to the new manner of engagement, such as auditor and auditees never meeting, or in some cases, not even being able to see each other during meetings where security settings did not permit the use of cameras on laptops. Of necessity, site visits were abandoned but document sharing, data analysis, and interviews were largely unaffected.

Like many knowledge organizations, the experience of audit staff working at home is leading to a reassessment of the use of remote working with potential financial savings and reduced carbon footprint being beneficial outcomes of new ways of working. The experience has shown it is possible

to audit remotely and to share information securely. Similar effects were felt by the NAO's parliamentary audience, with Parliament and its select committees (including the PAC) also working remotely or with hybrid arrangements (a mix of remote and physical attendance at hearings) which respected social distancing guidelines.

As well as affecting NAO working practices, an even more significant impact was on the work programme of the organization, with a shift to prioritize certain activities over others. The NAO website (National Audit Office, 2021) indicates that the pandemic led to a reassessment of its forward performance audit work programme. It stated:

> In the light of the wide range and significance of the government's actions to tackle the COVID-19 crisis, we are carrying out a broad and varied programme of work. We are looking at government preparedness for the pandemic, the spending on the direct health response and the wider emergency response. We are also looking at the measures aimed at protecting businesses and individuals from the economic impact. We will prioritise our work on areas where we think there have been particular challenges and where we feel there is most to learn.

More generally, the pandemic also had a similar effect on the NAO as on the rest of society of accelerating and accentuating many changes in the way it works. It has, for example:

- *Increased the pressure to report very quickly*; the first NAO reports on the government response to the pandemic were published within a few weeks of lockdown in May 2020; an investigation into the vaccines roll out was carried out in weeks, and was being updated just prior to publication to keep it current;
- *underlined the need for more risk-taking in reporting* because of the pace and the inadequacy of evidence trails, which has emphasized the importance of quality controls and rigorous assurance of the evidence base;
- *pushed NAO to diversify its outputs* with a regularly updated COVID "cost tracker" (National Audit Office, 2020), drawing together details of all government expenditure published in May 2020, and updated regularly thereafter so that a detailed, searchable record was available of what was being spent and by which government bodies; together with investigations, evaluative reports, good practice, and lessons learned outputs, all of which are discussed later; and
- *underlined even more the importance of objectivity and impartiality,* core audit attributes in a very contested world where the reputations of ministers and organizations are at stake in "real time," where there has been widespread concern at misinformation, over-promising and under-delivery by government, and where auditors have been particularly conscious of entering into highly sensitive spaces.

Canada

As in the UK's NAO, the OAG has experienced what it has described on its website as "far reaching impacts" in operations and made many adjustments to the way in which it works during the pandemic. As it noted:

> By mid-March [2020], OAG employees were working remotely, as were some or all the employees of organizations audited by the Office. All travel for OAG auditors had been suspended, in response to public health guidelines. Furthermore, the OAG had initial difficulties in obtaining audit evidence given the challenging work environment and reduced or closed access to the organizations the Office audits. As part of the Office's response to the pandemic, the OAG has collaborated with the federal organizations it audits, as well as with central agencies in the federal government, to adapt to the circumstances and develop workable solutions to complete its audit work.
>
> (Office of the Auditor General [OAG], 2020)

The OAG also identified several new risks related to the pandemic and financial reporting. These included changes in controls due to an increase of remote work and increased susceptibility to breaches in digital security, increased risk of fraud, and increased uncertainty in making some financial estimates.

In Canada, the OAG adjusted its financial work to meet all statutory obligations and delay other work. It has also made a number of significant, if less extensive, adjustments to its performance audit work because of the pandemic, albeit less extensive than those made by the NAO as its statutory framework, which provides for less flexibility in reporting than in the UK.

The OAG is established through the *Auditor General Act* and has a legislative basis in several other statutes which set out its powers and responsibilities. Legislation is needed to change these powers and responsibilities, such as the 1995 amendments that added a mandate related to environment and sustainable development, and reporting requirements on sustainable development strategies and environmental petitions. In addition to its established reports, the OAG on occasion produces other outputs such as the results of an auditor working group on climate change actions in 2018 or a study on establishing a First Nations Health Authority in 2015 (OAG, 2015, 2018).

The OAG meets with parliamentary committees, primarily the PAC, but also subject area committees and the Senate on the matters raised in its reports. However, whilst there have been innovations in the types of reports produced, the *Auditor General Act* limits the frequency of

reporting to Parliament. Thus, recent responses to the pandemic are made within that statutory framework and, as result, relate more to *how* the work of the Office has been carried out, changes in audit plans, and assessment of risk, rather than to *what* it does – unlike the diversification of outputs seen in the UK.

Like many other governments, the Canadian Government responded to the pandemic by establishing programs to support people who lost jobs, support businesses which were affected by the pandemic, to stabilize the economy, and slow the spread of the virus. This resulted in large, unanticipated expenditures. The public service did its part to deliver these initiatives and was asked to take on more risk. The Clerk of the Privy Council (the most senior public servant) in his 2020 Report challenged public servants to focus on delivering programs as quickly as possible, despite the inherent risk in this approach (Shugart, 2020). Public servants were encouraged to focus on urgent needs and achieving intended results, as well as to document decisions if actions were not aligned to the existing policy framework. In turn, the Auditor General acknowledged the challenges they were facing and the nimbleness of their responses. As she noted in her message to Parliament:

> The urgency and gravity of the pandemic pushed federal organizations in directions they might not have gone of their own accord, nor as quickly. This urgency and gravity provided federal organizations with the impetus needed to make a significant shift away from a process focus and toward a service mindset.
>
> (OAG, 2021a)

In spring 2020, the Canadian Parliament specifically requested the OAG to examine the government's pandemic spending and report to Parliament more quickly than normal. The OAG adjusted its work plan for performance audits by delaying audits planned for fall 2020 to spring 2021 and by deferring indefinitely audits planned for spring 2021 to respond immediately on pandemic spending. The Auditor General responds to motions of Parliament requesting an audit, but further decisions on the topics to audit are at his or her sole discretion. Given the urgency of efforts to slow the spread of the virus, the Auditor General took the unusual step of delaying audits such as on vaccine roll-out, which officials and the auditor felt could otherwise divert those involved from their work.

What has been produced in the UK

In the UK, the NAO's COVID-19-related knowledge production has fallen into five main areas. These are basic expenditure data, investigations,

evaluative reports, "lessons learned" products, and "good practice" products.

- *Gathering and publishing basic expenditure data:* Within a few weeks of the pandemic, the NAO published what it described a factual summary of the significant government spending commitments and programmes relating to COVID-19. The COVID-19 "cost tracker" (National Audit Office, 2020) is described as:

  > An interactive tool that brings together data from across the UK government. It provides estimates of the cost of measures announced in response to the coronavirus pandemic and how much the government has spent on these measures so far (where this information is publicly available or has been provided to us by central government departments).

  The purpose was described on the face of the "cost tracker" as to increase transparency and promote scrutiny and parliamentary accountability for government spending, core tasks of the NAO. The first was published in May 2020, with expanded and updated versions published in September 2020, January 2021, and May 2021. In April 2020, the NAO also stated the purpose of the cost tracker was to help identify a risk-based series of evaluative studies where we think there is most to learn. The data can be downloaded and analysed by type of support, department responsible and date of commitment.

- *Investigations:* As discussed in Lonsdale (2020, p. 126) investigations were designed by the NAO to be shorter than its performance audits and to be factual rather than evaluative. The reports set out facts in areas of public interest, rather than considering the value for money of government programs and were therefore seen as suited to the demands of reporting quickly to Parliament during the emergency. Investigations on the response to the COVID-19 pandemic were published from September 2020 onwards on, amongst other things, ventilator procurement, the "Bounce Back Loan Scheme," government procurement, preparations for vaccines, availability of free school meals, the housing of rough sleepers, support to charities, and the Culture Recovery Fund.
- *Evaluative reports:* The NAO has also looked, for example, at the impact of the pandemic on particular sectors, such as the need to reduce the backlog in criminal court cases, and on different groups, such as support for children's education during the early stages of the pandemic. These reports have taken a more evaluative approach by assessing performance and concluding on the effectiveness of interventions.
- *Lessons learned outputs:* Having published 17 reports in the year following the start of the pandemic, the NAO has also sought to draw out broader lessons and themes. These themes cover: risks to value

for money; transparency and public trust; data and evidence; coordination and delivery models; supporting and protecting people; and workforce and capability. Emerging lessons include the fact that UK was not well-prepared for a pandemic, with no contingency plans in place (Davies, 2021a). The head of the NAO also commented in the same article that there was:

> Much to be learned already from the pandemic. To promote transparency, government must clearly define its appetite and tolerance of risk, particularly under emergency spending conditions. Uncompetitive procurement practices must not be allowed to become a new norm. It should also monitor how Covid-19 programmes are operating, dynamically updating demand forecasts, and ensuring it has the ability to flex its response.

- *Practical guidance:* The NAO has also sought to draw on its on-going experience to support those in public bodies who are responsible for maintaining good governance in the more demanding and riskier environment. In March 2021, the NAO published a good practice guide covering financial reporting, the organizational control environment, and the regularity of expenditure (Comptroller and Auditor General 2020b). This was in response to concerns about the increased risk of fraud, error, and waste which were in part caused because some controls were no longer safe to operate, such as face-to-face meetings with applicants for benefits, or because there was a need to provide support to people and businesses quickly.

The variety of outputs reflects the flexible nature of the legislation governing the NAO and the conscious efforts that have been made in recent years to vary its publications for different purposes and audiences (Lonsdale, 2020). The urgent parliamentary demands for independently assured information meant that the NAO needed – and appears to have been able – to respond in different ways.

What has been produced in Canada

The OAG has continued to table its audit reports as they become available on priority topics in response to parliamentary requests. These reports differ depending on the topic being examined. For example, in March 2021, four reports were tabled, three of which dealt with COVID-19. These were:

Report 6: Canada Emergency Response Benefit

A benefit program of $500 per week for up to 28 weeks to support workers who lost income as a result of COVID-19 with a total

expenditure of about $74 billion. The audit examined the analysis and design of the program and whether it would support workers who lost their jobs. The audit found that government "considered and analysed key areas in the initial design and ongoing adjustments" (para 6.18) to include excluded workers. While more could have been done to introduce pre-payment controls, the Auditor General was sympathetic to the necessity of getting the program delivered quickly. About $500 million was identified as having been paid to ineligible recipients. The Auditor General will do a future audit once post payment controls have been implemented.

(OAG, 2021b)

Report 7: Canada Emergency Wage Subsidy

A subsidy program to help employers retain their employees and workers to maintain a source of income until June 2021. The program is expected to cost about $97.6 billion. As was the case for the audit of the Emergency Response Benefit, the focus of the audit was on appropriate up-front analysis and on whether appropriate controls had been put in place. The audit found that within a short time frame, a partial analysis had supported the program design followed by a sound and complete analysis to inform adjustments. The Auditor General recommended that a complete economic evaluation of the program be published. The audit also concluded that rapid implementation, the decision to avoid establishing tight controls, and gaps in business data resulted in not having all the necessary information, leading to having to "rely on costly comprehensive audits to recover payments made to ineligible recipients" (OAG, 2021c, para. 7.9). The data gaps identified existed prior to the pandemic. The Auditor General recommended strengthening efforts to increase compliance with the Goods and Services Tax and Harmonized Sales Tax that improve data gaps.

(OAG, 2021c)

Report 8: Pandemic Preparedness, Surveillance, and Border Control Measures

In Canada, health is a shared federal, provincial, and territorial government responsibility, adding to the complexity of health programming and surveillance. The audit examined whether the Public Health Agency of Canada was sufficiently prepared for the pandemic and whether their efforts were supported by good surveillance data. The audit also examined whether border controls were implemented and enforced. The audit found that the Public Health Agency of Canada was not well prepared to respond to the pandemic because long-standing surveillance issues (identified in past audits) had not

> been addressed; it had failed to regularly update or test plans for a pandemic response. Since the start of the pandemic, plans were further developed and work was undertaken to improve surveillance data, but outdated information technology issues still needed to be addressed. Borders were closed and incoming travellers were required to quarantine, but the Public Health Agency of Canada did not know whether two-thirds of incoming travellers actually followed quarantine orders. (OAG, 2021d)

In her opening statement to the PAC in April 2021 (Hogan, 2021), the Auditor General accepted that the decision to assume more risk in order to make timely payments to Canadians because of the emergency needed to be addressed. She also recommended that rigorous post-payment verification work be conducted, even though the government had already decided not to pursue $240 million of the Emergency Response Benefits which had been paid to ineligible self-employed individuals due to the personal hardships to individuals (The Canadian Press, 2021).

At the end of May 2021, two further reports (OAG, 2021e) were tabled. These were *Report 10: Securing Personal Protective Equipment and Medical Devices* and *Report 11: Health Resources of Indigenous Communities – Indigenous Services Canada*. The audits on the Government's response to the COVID-19 pandemic highlighted several long-standing, unaddressed problems including the maintenance of up-to-date pandemic plans, management of the National Emergency Strategic Stockpile, gaps in health surveillance data, and providing health care staff to Indigenous communities. They also pointed to known problems in having up-to-date information technology.

Challenges in knowledge production

The evidence of the first 18 months of undertaking performance audits and associated work in the context of the pandemic has shown that there have been a number of challenges in conducting this work. In particular:

- In many areas, basic data was challenging to collect or was not available at all. In the UK, the information from the first COVID cost tracker was confined largely to public domain material, with later editions expanded to audited data covering a wider range of activities.
- Gaining access to officials or operational sites in both countries was particularly difficult or impossible at times, particularly in the health sector, and auditors in both countries have limited their inquiries where appropriate. The auditors in both countries were keen not to disrupt the work of the bodies being audited. At the same time, it

was clear that some senior government officials recognized the value in a credible narrative being established and independent analysis being available, and so were often accommodating to the demands of audit.

- Normal criteria for judging performance were necessarily tempered by the realities of the situation. In the UK, the Comptroller and Auditor General commented that the NAO had "independently assessed each element of the Government's response based on what was reasonable to expect in the circumstances" (Davies, 2021a). Clearly consideration of what was "value for money" was very difficult and conclusions were caveated at times with recognition of the exceptional circumstances.
- The NAO commented, Our reports show how the trade-off between speed, effectiveness, cost, and control has been managed in the different elements of the COVID-19 response and provide important learning for the rest of this pandemic and any future public health emergencies.
- At the same time, expectations of appropriate conduct in public business remained, particularly given the scale of spending. As a result, concerns have been expressed in the UK, for example, about a lack of transparency around key decisions, particularly where suppliers were chosen for large contracts involving billions of pounds.
- The speed of data capture and weakness in some of the evidence base has been a challenge to normal ways of working. Particular attention has therefore been paid by auditors to their own internal quality checks, often in real time alongside data gathering.

How have auditors been able to respond in these ways?

The overall impression in the UK and Canada is that the performance auditor response has increased in speed, agility, and engaged in more risk-taking than normal whilst avoiding interfering with the activities of government. The NAO stated (National Audit Office, 2021) that:

> Our challenge is to try and provide the appropriate level of evidence-based reporting to support accountability and provide insight at the most suitable time. We must not get in the way of public servants working hard to save lives, but we must also ensure that our reporting is sufficiently prompt to support proper accountability for public money.

Auditors have been able to respond in the ways they have because they believed that this was what was expected of them. From the outset of the

pandemic, the NAO in the UK considered there were expectations about what its role should be. It stated on its website in summer 2020:

> What is already clear is that MPs, and the public that they represent, will expect us to carry out a substantial programme of work on the COVID-19 response so we can learn for the future. This will include looking at government spending on the direct health response as well as the wider emergency response. We will also look at the spending on the measures to protect businesses and individuals from the economic impact.

As well, auditors in both countries have an existing presence across government and therefore knowledge of the organizations and programs that came under enormous pressure, as well as existing networks and contacts, which allowed them to carry forward their work looking at exceptional issues as part of their normal responsibilities.

The UK's NAO has discretion over its work program, so it could pivot to address the new demands placed on the government, without waiting to be commissioned. It could also free up resources by dropping other, less pressing work and change direction whilst remaining entirely within its remit. As discussed in Lonsdale (2020), it had already been shifting in the direction of faster, more diverse reporting, and therefore the need for rapid, often fact-based examinations was in keeping with its existing trend.

Finally, both SAIs were strongly supported by parliamentarians who have been very keen to have access to objective evidence and advice. The OAG has also responded to parliamentary interests by adjusting its plans, adapting its view of risk, and not interfering in government operations. It has not, however, shifted to more diverse reporting because of the nature of its governing legislation. It relies primarily on its performance audits to provide information on government expenditure, programs, and their results. It has produced fewer outputs but has responded to the parliamentary requests to examine pandemic spending. The Auditor General appeared before the PAC to discuss these reports.

In Canada, the OAG experienced similar pressures to the NAO to respond to the pandemic but did so at a time when it was also going through a number of internal changes in leadership, including the appointment of a new Auditor General. The organization also considered that long-standing resource pressures affected how easily it was able to respond. In May 2020, the Interim Auditor General told Parliament that lack of funds had left it with no choice but to delay work on most audits as the COVID-19 pandemic added new demands on what was described as the "resource-stretched office" (Lim, 2020). This meant that resources were diverted to COVID-19-related work

and the timetables for existing work not related to motions adopted by the House of Commons were revisited.

Parliamentary consideration of COVID-19 expenditure

Performance audits are often used by legislatures to hold government to account (Lonsdale et al., 2011). The NAO's original work was primarily to establish facts and provide analysis, and the House of Commons' PAC has held public sessions on each of the individual topics covered by the NAO's COVID-19 reports, published at different points in the year, taking evidence from the officials responsible and issuing its own reports (see, e.g., Public Accounts Committee, 2021).

The OAG's relationship with Parliament is based on the British "Westminster model" but has evolved some different practices. The Auditor General tables her reports with observations, conclusions, and recommendations that are referred to the House of Commons PAC. The PAC holds a hearing with witnesses that include the Auditor General and issues a report of its deliberations. Unlike in the UK, the Auditor General normally submits two reports to Parliament on several performance audits together each year, one in the spring, the other in the fall.

Performance audits allowed parliamentarians to consider topics quickly, establish a basic understanding of issues, and make recommendations for action whilst the pandemic was still under way and lessons could be applied to handling subsequent stages. The traditional cross-party nature of the Parliamentary PAC in the UK has helped to diffuse the risk of a partisan analysis which has not always been the case in Canada.

Conclusions

The two SAIs covered in this chapter are well respected and established national organizations. The responsiveness of their performance auditors and the flexibility of their work have been crucial to its usefulness in helping the Parliaments of both countries to continue to scrutinize the use of enormous sums of public money – much of it unplanned and spent rapidly during the pandemic – and to maintain some measure of oversight in exceptional circumstances. In different ways, they have adjusted their work and audit plans in response to the COVID-19 pandemic. Such adjustments have, by necessity, been made within existing statutory and policy frameworks. The permissive and flexible legislation within which the NAO operates in the UK has enabled it to effectively use different forms of reporting, tailored to the needs of the time. The OAG in Canada has been more constrained by its legislation and resource constraints but has nevertheless tilted its work program to the demands of the day.

In order to maintain relevance and utility, both audit bodies have also exploited their respective existing understanding of the organizations delivering services and used their long-standing experience and powers of access within accountability regimes. This has enabled them to provide the objective knowledge, data, and insight that elected representatives have sought in times of crisis. These case studies from the UK and Canada have highlighted the crucial importance of timeliness and flexibility in knowledge production in order to be useful and valued in this time of emergency. The OAG has addressed Parliament's requests within its traditional reports but has amended its work program to accommodate them. The NAO has also demonstrated that very fast preparation of reports – even necessitating some compromises – was acceptable if there was transparency about the limitations of the methodology, including, for example, the restrictions on data collection or inability to make site visits. Both SAIs have faced and responded to the challenge of adjusting traditional ways of viewing and assessing risk in public service delivery to make realistic judgements about exceptional levels of spending.

The traditional strengths of SAIs – independence, understanding the workings of government, the ability to determine their own work programmes and report publicly – have allowed them to adapt to differing degrees and in different ways to the challenges of the pandemic. Both have had to quickly learn to work differently and are likely to integrate lessons from the experience of remote working such as reduced travel costs, increased virtual contact, possibly expanding the range and representativeness of participants in audit meetings, and flexible home working for staff into their standard operating models in the future. At the same time, they will seek to avoid some of the potential risks, such as loss of personal contact with those delivering services, reduced opportunities for informal knowledge gathering and seeing activities on the ground, and the undermining of organizational cohesion because of staff spending less time together. The importance of traditional audit concerns – the regularity and propriety of public spending, as well as its effectiveness, and the identification and management of risk – has only been reinforced by the experience of the pandemic, including public reinforcement of the value of the role of independent reporting on these matters.

At a higher level there are, and will continue to be, many questions to be considered about the effectiveness of government programs and spending. Some of these will be examined within government, but others will be much more visible through public inquiries. Both forms of inquiry will ask similar questions. Should governments have been better prepared? How can they prevent the worst effects of such an emergency in future has been required to adapt? How effective have government interventions been and why? State audit bodies – as demonstrated by the work of the NAO and OAG undertaken since March 2020 – are well positioned to contribute knowledge and learning generated during the pandemic from their performance audits to

both the internal and external debates which will inevitably take place. Their performance audit reports, representing trusted information based on the best data available, may offer important insights for these inquiries on many of the key questions. SAIs will thus be able to contribute to learning and help strengthen the resilience of government in the future.

At the same time, governments in power leading the fight against the pandemic will be held accountable for their decisions and for the effectiveness of their responses. Notwithstanding the often-bipartisan nature of much of the discussion during the period of crisis (with political rivals tending to rally behind those in power during the periods of peak emergency), opposition parties and government critics are likely to provide increasing challenge once the pandemic is over, leading to far more partisan debate about what happened and when, why particular choices were made, and how effective were government interventions. Here again, SAIs – with their key place in accountability processes – will have an important role to play in providing the evidence on which government bodies and ministers will be held to account. The nature of what auditors have examined and reported on, such as the support programmes for individuals and businesses (including the differential impact on different groups), preparedness measures, and the roll-out of health measures will help to ensure that the issue of concern to citizens is central to public scrutiny.

This underlines the fact that the contribution to the public debate on the impact of the pandemic of those auditing and evaluating may well depend on where they work. Internal government evaluators will inevitably be part of the government response, and so do not have the independence needed to judge effectiveness. In addition, those evaluation bodies which have been constrained in their ability to work by the effects of the pandemic, for example, by restricted access to subject organizations, may have fewer insights to offer. In contrast, those independent bodies, such as SAIs, which have had a "ring side seat" during the challenging times, will be well placed to contribute through their analysis and knowledge to the crucial assessments of what happened during the management of the pandemic and how things could be done better.

References

Barrados, M., & Lonsdale, J. (Eds.). (2020). *Crossover of audit and evaluation practices*. Routledge.

Comptroller and Auditor General (2020a, June 22). *Guide for audit and risk committees on financial reporting and management during COVID-19*. National Audit Office. https://www.nao.org.uk/wp-content/uploads/2020/06/Guide-for-audit-and-risk-committees-on-financial-reporting-and-management-during-Covid-19-1.pdf

Comptroller and Auditor General (2021, June 21). *Test and trace in England – Progress update*. National Audit Office. https://www.nao.org.uk/wp-content/uploads/2021/06/Test-and-trace-in-England-progress-update.pdf

Davies, G. (2020b, November 13). *Investigation into government procurement during the COVID-19 pandemic*. National Audit Office. https://www.nao.org.uk/wp-content/uploads/2020/11/Investigation-into-government-procurement-during-the-COVID-19-pandemic.pdf

Davies, G. (2021a, April 22). *Auditing government's pandemic response*. NAO Blog. https://www.nao.org.uk/naoblog/auditing-governments-pandemic-response/

De Bruijn, H. (2007). *Managing performance in the public sector* (2nd ed.). Routledge.

Hogan, K. (2021, March). *Message from the Auditor General of Canada*. Office of the Auditor General of Canada. https://www.oag-bvg.gc.ca/internet/English/osh_20210413_e_43800.html

Lim, J. (2020). *Nominee for Auditor General seeking permanent stable funding as a first priority*. iPolitics. https://ipolitics.ca/2020/05/19/nominee-for-auditor-general-seeking-permanent-stable-funding-as-a-first-priority/

Lonsdale, J. (2020). Auditing in changing times: The UK National Audit Office's response to a turbulent environment. In M. Barrados, & J. Lonsdale (Eds.). *Crossover of audit and evaluation practices*. Routledge.

Lonsdale, J., & Mayne, J. (2005). "Neat and tidy…and 100% correct" assuring the quality of supreme audit institution performance audit work. In R. Schwartz, & J. Mayne (Eds.), *Quality matters: Seeking confidence in evaluating, auditing, and performance reporting* (pp. 173–198). Transaction.

Lonsdale, J., Wilkins, P., & Ling, T. (2011). *Performance auditing: Contributing to accountability in democratic government*. Edward Elgar.

Marrocco, F. N., Coke, A., & Kitts, J. (2021). *Ontario's long-term care COVID-19 commission final report*. Government of Ontario. http://www.ltccommission-commissionsld.ca/report/index.html

Mayne, J. (2007). Evaluation for accountability: Myth or reality. In M.-L. Bemelmans-Videc, J. Lonsdale, & B. Perrin (Eds.). *Making accountability work: Dilemmas for evaluation and for audit*. Transaction.

National Audit Office. (2020). *COVID-19 cost tracker*. www.nao.org.uk/covid-19/cost-tracker/

National Audit Office. (2021). *Supporting Parliament to understand the government's response to the coronavirus pandemic: Our programme of work*. www.nao.org.uk/covid-19/

Office of the Auditor General. (2015). *Establishing the First Nations health authority in British Columbia*. https://www.oag-bvg.gc.ca/internet/English/parl_oag_201602_07_e_41064.html#hd3b

Office of the Auditor General. (2018). *Perspective on climate change action in Canada – A collaborative report from Auditors General – March 2018*. https://www.oag-bvg.gc.ca/internet/English/parl_otp_201803_e_42883.html

Office of the Auditor General. (2020). *COVID-19 impact on the OAG*. https://www.oag-bvg.gc.ca/internet/English/parl_fs_e_43616.html

Office of the Auditor General. (2021a). *Message from the Auditor General March 2021 to Parliament*. https://www.oag-bvg.gc.ca/internet/English/parl_oag_202103_00_e_43782.html

Office of the Auditor General. (2021b). *Report 6 – Canada emergency benefit*. https://www.oag-bvg.gc.ca/internet/English/parl_oag_202103_01_e_43783.html

Office of the Auditor General. (2021c). *Report 7 – Canada emergency wage subsidy*. https://www.oag-bvg.gc.ca/internet/English/parl_oag_202103_02_e_43784.html

Office of the Auditor General. (2021d). *Report 8 – Pandemic preparedness, surveillance, and border control measures*. https://www.oag-bvg.gc.ca/internet/English/parl_oag_202103_03_e_43785.html

Office of the Auditor General. (2021e). *2021 reports 10 and 11 of the Auditor General of Canada*. https://www.oag-bvg.gc.ca/internet/English/parl_oag_202105_e_43832.html

Public Accounts Committee. (2021). *Initial lessons from the government's response to the COVID-19 pandemic*. UK Parliament. https://committees.parliament.uk/work/1225/initial-lessons-from-the-governments-response-to-the-covid19-pandemic

Shugart, I. (2020, July 21). *27th Annual report to the Prime Minister on the state on the public service of Canada*. Government of Canada. https://www.canada.ca/en/privy-council/corporate/transparency/annual-report-prime-minister-public-service/27th.html

The Canadian Press. (2021). *30,00 self-employed people may not have to repay CERB*. CP24. https://www.cp24.com/news/cra-estimates-30-000-self-employed-people-may-not-have-to-repay-cerb-1.5445673?cache=yes%23responsive_main%23responsive_main%2F5-things-to-know-for-monday-october-28-2019-1.4658248

5 The Impact of the COVID-19 Pandemic on the Effective Use of Evaluation in Supporting the Sustainable Development Goals

Robert Lahey and Dorothy Lucks

Introduction

The 2030 Agenda for Sustainable Development (the "2030 Agenda"),[1] presents a unique opportunity to positively shape how societies grow and develop based on 17 Sustainable Development Goals (SDGs). Initially, measuring progress toward the SDGs was focused on the use of indicators and targets. In many countries, evaluation was not a high priority. However, global stakeholders[2] stressed that evaluation is "a crucial ingredient for SDG success" and highlighted the need to build national evaluation systems for countries in meeting their commitment to and achievements in SDG implementation, management, monitoring, evaluation, and reporting.

The COVID-19 pandemic has affected this trajectory. Expectations set before the pandemic were overtaken by the massive emergency response and resource reallocation required to address immediate public health needs and ongoing economic recovery (Grossi et al., 2020). This places an imperative on the evaluation sector to consider how evaluation theory and practice should respond to the effects of the pandemic to ensure ongoing relevance, effectiveness, and efficiency in supporting SDG achievement.

The purpose of this chapter is to explore two key questions: (i) What has COVID-19 meant for the use of evaluation in supporting the implementation, management, and reporting on the SDGs and (ii) how can evaluation practitioners adapt in order to remain relevant to the SDGs and country-level needs going forward?

The chapter provides a brief background to monitoring, evaluation, and learning (MEL) in the context of the SDGs, with a focus on evaluation practice as a critical contribution to their progress. It introduces an SDG MEL framework, then applies this framework to highlight where and how COVID-19 is affecting the use of evaluation in supporting the SDGs at the country level, how evaluation has responded to date, and what the most important responses should be going forward. Information to develop this paper was drawn from a wide variety of sources, including recent documentation regarding COVID-19, evaluation practice and the SDGs across

DOI: 10.4324/9781003376316-6

different countries, perspectives of select United Nations (UN) and international development agencies in support to national development, as well as the experience and assessment of evaluators and researchers from around the world.

Based on analyses of these sources, the paper provides lessons learned regarding the role and practice of evaluation to ensure the continued relevance and increasing usefulness of evaluation in relation to the 2030 Agenda.

Background – Pre-pandemic strengths and weaknesses in the use of evaluation with the SDGs

The 2030 Agenda establishes 17 SDGs, with 169 agreed targets to be achieved by the year 2030.[3] Each country is responsible for its own progress toward the SDGs and their targets. The targets were designed to unite efforts toward sustainable outcomes, provide coherence in measuring progress and inspire active and integrated problem-solving toward a more prosperous and resilient future (Sustainable Development Solutions Network Secretariat, 2014). The 2030 Agenda outlines several key aspects, including: (i) integration/coherence; (ii) "leaving no one behind;" (iii) balanced action on all three elements of sustainable development; (iv) equity; (v) resilience; (vi) universality; and (vii) mutual accountability, all of which are intended to guide all sustainable development initiatives (G.A. Res. 70/1, 2015).

A key observation in the transitioning from the Millennium Development Goals (MDGs) to the SDGs was the insufficient attention paid to monitoring performance and analyzing results (United Nations Development Programme [UNDP] & World Bank, 2016). The 2030 Agenda stresses the importance of accountability for progress and calls for countries to engage in a systematic "follow-up and review" of their work to achieve the SDGs through country-led voluntary national reviews (VNRs) (UNDP & World Bank, 2016). Two key mechanisms at the country level are intended to measure progress and analyze performance: (i) global indicators linking to specific SDGs and (ii) country-led evaluation of SDG progress and performance, reported through the VNR process. Evaluating SDG progress would focus on identifying achievements and enhancing learning and innovation. The High-Level Political Forum (HLPF) acts as a platform for all countries to showcase their performance by presenting their national reviews and policy discussion on selected SDGs.

Awareness has over time increased for global and national leaders on the use of evaluation and its importance to the SDGs. This has developed through the efforts of a variety of international players and networks – EvalPartners, EVALSDGs, the International Organization for Cooperation in Evaluation (IOCE), UN Evaluation Group (UNEG) and its various

UN members, as well as EvalNet and, more recently, Eval4Action and the Global Evaluation Initiative.

Some evidence of the improvements achieved to date comes via the VNRs reporting on progress toward the SDGs prior to 2020.[4] There is increased incidence of planning to use evaluation for analyzing and reporting on SDG progress. Countries such as Botswana and Costa Rica, among others, have developed an SDG roadmap that speaks not only to implementation, but also to the monitoring, evaluation, and reporting on performance and progress of the SDGs, which were considered by country officials to be crucial elements in informing and integrating into public policy decisions (Government of Costa Rica, 2018; Matambo et al., 2018). In broader terms, the Global Partnership for Effective Development Cooperation (GPEDC) data in 2018 and 2019 shows that 64 percent of countries have high quality national development strategies in place (Organisation for Economic Co-operation and Development [OECD] & UNDP, 2019).

Global evaluation leaders and international partners such as the UNEG have also been working with national stakeholders to support national evaluation capacity development so as to link country-led evaluation to the SDGs (United Nations Evaluation Group [UNEG], 2015). But, while there have been reports of strengthening national monitoring and evaluation system (NMES) initiatives, the same GPEDC data reports that only 35 percent of countries had monitoring and evaluation (M&E) systems in place to track the progress of national strategy implementation (OECD & UNDP, 2019).

The availability of capacity development, guidelines, and implementation has increased. For example, a guide entitled *Evaluation to connect national priorities with the SDGs* was disseminated to national governments and accompanied by a series of webinars.[5] There have been improvements in data development, a crucial ingredient for both M&E. For instance, in Cuba, Norway, and Zimbabwe, parliamentary processes have mandated analyzing implementation of the 2030 Agenda as part of their M&E work through mainstreamed commission or committee functions. In Indonesia, governors coordinate and report SDG implementation to the Minister of National Development and the Minister of Home Affairs. A Sub-National Coordination Team strengthens the involvement and the role of all stakeholders including non-state actors from philanthropy and business, academia, and civil society organizations (United Nations Department of Economic and Social Affairs [UNDESA], 2021).

Yet, most VNRs note an ongoing challenge with data limitations that continue to constrain SDG monitoring and evidence-based planning (UNDESA, 2019). Major gaps in national M&E systems remain. Significant efforts are needed at the country level to build both capacity and a suitable enabling environment to facilitate both the conduct and "use" of evaluation. This in turn requires further commitment by leadership to develop and resource the evaluation capacity needed for effective SDG MEL.

If evaluation is indeed "a crucial ingredient for SDG success," questions arise as to where and how evaluation needs to adapt in order to be relevant and useful in supporting the SDGs in this changing global and national context. The application of a SDGs MEL Framework could provide a systematic approach to identifying and interpreting the trajectory of progress for application of evaluation for the SDGs and the 2030 Agenda, as well as identifying how the evaluation sector can adapt to the effects of COVID-19.

An SDG monitoring, evaluation, and learning (MEL) analytical framework

There is no "one-size-fits all" approach in national responses to the SDGs or to COVID-19 (Hale et al., 2022). There is limited "hard" data describing the impact of the pandemic on the role and use of evaluation in supporting the SDGs at the country level. A lack of data also relates to the absence of a MEL framework for the COVID-19 response. This means that efforts to address the socio-economic impact effectively are not being sufficiently and systematically tracked, or assessed to identify what works, what does not work, and what is likely to be faced in the future.

The framework shown in Figure 5.1 was devised to serve as a mechanism for identifying where and how the COVID-19 pandemic may be impacting the use of evaluation in supporting a country's implementation of the SDGs. It identifies success factors underlying the effective use of evaluation in supporting the SDGs. The framework was drawn from two earlier sources: (i) previous analyses addressing the evaluability of the SDGs[6] and (ii) analysis of the underlying factors and conditions for an effective NMES, including a framework for evaluation capacity building at the country level.[7]

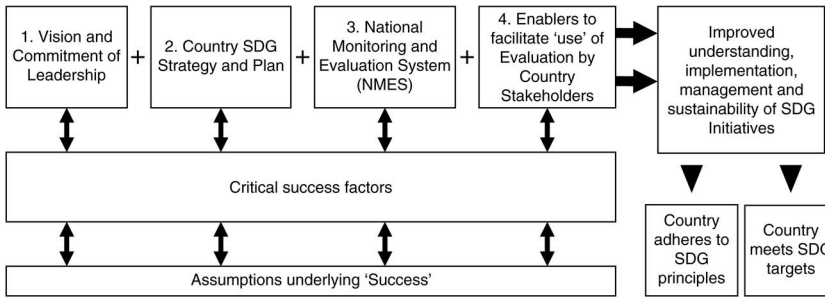


Figure 5.1 SDGs MEL Analytical Framework.

Source: Robert Lahey

The four key elements of the framework providing a broad frame of reference represent factors underlying the effective use of evaluation in supporting the SDGs: leadership; SDG strategy and plans; NMES; and enablers for use. Together these factors contribute to improved understanding, implementation, management, and sustainability of SDG initiatives and ultimately to supporting SDG principles and meeting SDG targets.

Critical success factors, shown in Table 5.1, give form and direction for how a country could position evaluation and the SDGs for effective use of evaluation to support action for SDG achievement. The "critical success factors" are based on expectations regarding country commitments to SDG implementation, and good practices regarding evaluation development and use in supporting good governance at the country level. The framework emphasizes that, in analysis, underlying assumptions for the critical success factors to operate effectively also need to be considered and tested. Such assumptions are shown in Table 5.2.

Table 5.1 Critical success factors – Effective evaluation for SDG progress.

Key Element	*Critical Success Factors*
1. Vision and Commitment of Leadership	1.1 Priority given to the SDGs 1.2 Appreciation for the importance and role of evaluation 1.3 Political support and demonstrated commitment to evaluation 1.4 Political support and demonstrated commitment to public sector adaptation and improvement
2. Country SDG Strategy and Plan	2.1 SDG planning 2.2 SDG coordination and implementation 2.3 Systematic monitoring of the SDGs 2.4. Identified priorities for evaluation of SDG-related initiatives
3. National Monitoring and Evaluation System (NMES)	3.1 Institutional structure supporting monitoring and evaluation 3.2 Evaluation policies, guidelines, and practice standards 3.3 National and sub-national evaluation capacity 3.4 Maturity of evaluation practice
4. Enablers to Facilitate "Use" of Evaluation by Country Stakeholders	4.1 Drivers and uses of monitoring and evaluation information 4.2 National-level facilitators to coordinate development and use of monitoring and evaluation 4.3 Organization-level facilitators activating and supervising use 4.4 Active networks and partnerships to facilitate learning and awareness

Table 5.2 COVID-19 impact on MEL success factors and assumptions for the SDGs.

Critical Success Factor	*Underlying Assumptions for Success*	*COVID-19 Impact*
Vision and Commitment of Leadership		
1.1 Priority given to the SDGs.	Leadership of the country recognize the importance of and give priority to implementing the SDGs.	Changing national priorities.
1.2 Appreciation for the importance and role of evaluation.	Awareness and understanding that information from ongoing monitoring and periodic evaluation can assist public sector managers, decision-makers, and the country in moving to achieve national goals.	Continuous.
	Evaluation is recognized as a tool for both *learning* and *accountability*.	Continuous.
1.3 Political support and demonstrated commitment to evaluation.	Political support for resourcing and use of systematic evaluation as part of the governance and decision-making structure in the country.	Reallocated and stretched resources.
	Transparency and accessibility of M&E information to the media and civil society for their participation in the national monitoring and evaluation system.	Changes in data collection, distribution, and reporting. Some questions on transparency.
1.4 Political support and demonstrated commitment to public sector renewal and improvement.	Commitment to public sector improvement and, where needed, public sector reform.	Reform attention shifted.
	Willingness to challenge the status quo and current culture within organizations.	Status quo and culture shifting; uncertainty of the future.
	Leadership encourages and fosters initiatives aimed at improving accountability.	Continuous.
Country SDG Strategy and Plan		
2.1 SDG planning.	The SDGs are linked to national priorities.	Changing national priorities.
	National SDG strategy and detailed plan for implementing the SDGs.	Changing priorities and resources impacting SDG strategy and plans.

(*Continued*)

Table 5.2 COVID-19 impact on MEL success factors and assumptions for the SDGs. (*Continued*)

Critical Success Factor	*Underlying Assumptions for Success*	*COVID-19 Impact*
2.2 SDG coordination and implementation.	Coordination of implementation of SDG initiatives across agencies.	Coordination focus changed.
	All sectors – public, private, and civil society – are engaged in the development and roll-out of initiatives aimed at achieving SDG targets.	Attention shifted to COVID-19 response.
2.3 Systematic evaluation of the SDGs.	The SDG plan includes milestones for systematic evaluation, reflection on the results and lessons learned to that point, and course changes as needed.	Timelines delayed or changed.
National Monitoring and Evaluation System (NMES)		
3.1 Institutional structure supporting MEL.	Institutional structure that supports the ongoing collection of performance information and the periodic conduct of systematic evaluation.	Continuous.
3.2 Evaluation capacity.	Sufficient national evaluation capacity – resources and skilled personnel with technical capacity, competencies, and experience with evaluation – to conduct or manage evaluations.	National evaluation capacity required; changing needs from experienced evaluators.
	Resourced capacity for ongoing training, development and skills upgrading of M&E practitioners.	Resource allocation shifted.
3.3 Maturity of evaluation practice.	Regular conduct and use of systematic evaluation of projects, programs and policies that assesses their delivery, effectiveness, and continued rationale.	Regular timelines disrupted; needs from evaluators changing.
	The conduct of evaluation engages all sectors – public, private, and civil society.	New challenges for stakeholder engagement.
3.4 Evaluation policies, guidelines, and standards.	The practice of evaluation within the country reflects professional practice standards and ethical guidelines of international "good practices."	Challenges for effective supervision of MEL quality; changes to evaluation practice.

(*Continued*)

Table 5.2 COVID-19 impact on MEL success factors and assumptions for the SDGs. (*Continued*)

Critical Success Factor	*Underlying Assumptions for Success*	*COVID-19 Impact*
Enablers to facilitate "use" of evaluation by country stakeholders		
4.1 Drivers and uses of monitoring and evaluation information.	Clarity around where and how MEL information is used at a project, program, ministry/sector, and national levels.	Changing needs for where and how evaluator's skill set is usefully employed.
	Formal guidelines or policies exist to inform and instruct officials on where and how M&E information is to be used, presented, and reported within the public sector.	Emerging needs go beyond traditional guidelines for evaluation.
	MEL is part of the normal process linked with discussions and decisions related to program development, policy, planning, budgeting, and reporting.	Continuous.
4.2 National-level facilitators.	Coordination of systematic national M&E efforts across the national statistical agency, agencies of government, and all other measurement and reporting efforts.	New challenges for coordination processes.
	A data development strategy and action plan exists for the country to address data deficiencies, including gaps in sub-national and demographic data needed for "results"-oriented analysis.	Focus on and resources for data processes shifted; greater demand for sub-national data.
4.3 Organization-level facilitators.	Institutional capacity within organizations to "use" M&E information. For example, forums exist that serve as mechanisms for reporting, sharing, and using evaluation results.	Internal mechanisms and procedures altered during period of emergency.
	An accountability for using evaluation information is established within organizations, with monitoring of how and how well information is being used.	Existing and emerging performance management systems challenged.

(*Continued*)

Table 5.2 COVID-19 impact on MEL success factors and assumptions for the SDGs. (*Continued*)

Critical Success Factor	*Underlying Assumptions for Success*	*COVID-19 Impact*
4.4 Active networks and partnerships.	An active professional network exists to support information exchanges and training, development and awareness-raising of evaluation concepts, methods, and practices.	In-person training and development not feasible; emergence of virtual events.
	Civil society, the private sector and the media are adequately informed about the role that monitoring and evaluation can play in good governance.	Shift in focus; uncertainty around MEL role in pandemic.

Impact of the COVID-19 pandemic on progress for SDGs and MEL

COVID-19 has drawn attention away from the fundamental developmental challenges and the initial momentum generated through SDG implementation. The 2021 HLPF resulted in a statement of renewed commitment to the 2030 Agenda and focused on the theme of "Sustainable and resilient recovery from the COVID-19 pandemic [...]."[8] The worldwide focus is to address the imperative of pandemic response and recovery; this is leading to a need in the shift or re-balancing of priorities. Research on the effects of COVID-19 demonstrates the substantial impacts that have resulted in significant socio-economic losses and negative progress toward SDG targets (Abidoye et al., 2021). As part of the UN response to COVID-19, a UN Framework for the immediate socio-economic response to COVID-19 was prepared with ten key performance indicators (United Nations [UN], 2020a). This led to mobilization of UN Country Teams to collaborate with national leaders to develop national UN socio-economic response and recovery plans. These plans have assisted in the mobilization of resources toward critical priorities for COVID-19. Key indicators were proposed with detailed indicators and a review process expected to form part of the national plans, but there is no integrated monitoring, evaluation, or reporting process. The link to and effect on SDG implementation are not an explicit factor of these efforts.

The pandemic has created a much higher profile for the use of science and "evidence" in supporting the decisions and actions of senior officials (Anessi-Pessina et al., 2020). The demand for national statistics is high as governments and other major stakeholders make major decisions. National statistical agencies continued to function despite major impediments to data

capture. The Organisation for Economic Co-operation and Development found that the COVID-19 pandemic had a significant impact on the compilation and dissemination of official statistics: lockdowns in many countries and teleworking affected how surveys and censuses were carried out.[9] At the same time, there is huge user demand for early estimates to enable assessment of the economic and social impacts of the crisis (Committee on Statistics and Statistical Policy, 2020).

For all countries, COVID-19 brings new attention to the SDG principle of "leave no one behind." An immediate challenge created by the pandemic is dealing with the health, economic, and social impact on the most vulnerable, a group that has been growing larger throughout the pandemic as more people experience heightened vulnerability due to social conditions driven by COVID-19, such as prolonged confinement and fracturing of social support networks. It has been recognized that the pandemic has globally widened the gap in terms of the disadvantaged groups being left behind (Florida, 2020). In tracking the unequal nature of the COVID-19 crisis, data shows that women have been impacted most by the crisis, and there is every likelihood that the recovery will be equally discriminatory (UN, 2020b). Aspects of social distancing are becoming embedded into social practice, with remote, digital meetings becoming prevalent. Electronic communication has become more normalized and different forms of exclusion are likely to emerge as health protocols also shift. Isolation is an integral aspect of marginalization and evaluators need to adjust to new methods of consultation and data gathering.

Challenges that existed prior to the pandemic (such as structural challenges) remain, and in some cases have worsened, but leadership in many countries cite the desire to "build back better" after the pandemic. It is likely that countries will review and "re-set" their SDG strategy and plan, as well as revisit the appropriateness of earlier SDG targets. Of importance is a better understanding of the impact the pandemic has had on various groups in society. This heightens the need for data at a sub-national level and disaggregated demographic analysis, areas which generally lag in terms of data development in many countries.

At the country level, in some cases, national SDG strategies, plans, and systems are being modified during the pandemic as country priorities are changing. A country's SDG plan may not have altered in the short-term, but there is every likelihood that events will force some modification of a medium-term national development plan (UN, 2020a). For instance, in Fiji, the government has used the pandemic period to plan and install energy-saving measures in tourism infrastructure and to push forward on a climate change bill.[10] This shift in policy is measurable and traceable through the SDG indicators, other global and national indicators, and through consultation with key stakeholders. Countries have grappled with the multiple constraints of addressing the pandemic but there are lessons

that can be drawn from evolving practice in relation to the SDGs' implementation and evaluation.

Applying the SDG MEL framework demonstrates COVID-19 impact on evaluation and the SDGs

Using the SDGs MEL Framework, aspects of change due to COVID-19 and the impact on evaluation and the SDGs are far-reaching. A detailed analysis is shown in Table 5.2; applying the SDGs MEL framework shows how all of the elements and most of the success factors are affected by the pandemic's impact with only a few critical success factors that are continuous and unimpeded by COVID-19 as they are already embedded in on-going evaluation practice. Testing the framework assumptions for success identified potential issues that the evaluation sector needs to address.

The evaluation sector response to the COVID-19 pandemic

There is considerable disruption for the evaluation sector related to the SDGs. The pandemic highlighted the importance of national data systems for effective COVID-19 reporting and contact tracing. Health data systems have been activated in an unprecedented manner. Greater focus is being placed on monitoring some key health-related indicators and mathematical modeling. However, less focus is being given by decision-makers to more in-depth social science approaches of evaluation. In terms of "drivers" for evaluation, there have already been many calls within countries to carry out a post-mortem on their COVID-19 response. In so doing, long-term perspectives to help inform policy thinking and priority-setting need to draw on interdisciplinary approaches. These changes require systematic review, research, and consultation, areas where evaluation can provide the necessary skill set to country leaders. If evaluation is to be effective and sustainable at the country level, the experience of the pandemic period has reinforced the need to bring evaluation closer to decision-makers in a timely fashion.

For countries needing investment in MEL capacity, resources may be earmarked for what are deemed higher immediate priorities. Implications that are likely to affect implementation of MEL in the future include continuing uncertainty, fiscal challenges, reduced implementation capacity, and socio-cultural changes. Future resourcing of MEL capacity-building could be at risk, as there is generally no clear direction on how budgets or debt associated with the massive expenditures at national and sub-national levels will be financed going forward. MEL practice needs to respond rapidly to address the emerging issues and opportunities.

A "value-added" role and approach for evaluation: An evolving context

This role could include a focus on the SDG principles rather than the specific indicators and targets and take a more context-specific approach. For example, during the pandemic, clustered evaluations have become not just a policy imperative but also the preferred modality, as development partners appreciate the need for focused and strategic evaluations with a reduced burden on constituents and other stakeholders. Traditional evaluation approaches are less relevant for short-term priorities, but the evaluators' analytical skill set can be highly valuable. Evaluation carried out at the country level has the potential to be incisive in assessing both the pandemic response and recovery, and the trajectory of SDG progress. But the extent to which it can, or is currently, achieving these assessments is not clear.

The urgency of the pandemic and the importance placed on "just-in-time" information for decision-making will impact expectations about timely delivery of evaluation, an area often criticized in the past. The use during the pandemic of rapid consulting-type services using innovative and rapid approaches to evaluation raises the need to re-examine professional practice standards and what are deemed to be international "good practices" for evaluation. Collaborative models of evaluation, such as the Multilateral Organisation Performance Assessment Network's (2021) thematic assessment *Pulling Together: The Multilateral Response to Climate Change*, could provide a means of meta-evaluation that can accelerate learning through evaluation in a way that can be of use to multiple audiences.

This elevates the importance of international initiatives to support countries needing to build evaluation capacity, such as the Global Evaluation Initiative of the World Bank, UNDP, and UNEG initiatives. It will be important to re-energize global MEL efforts and to work toward institutionalizing evaluation in countries where there is limited formal presence. International partners and agencies are engaging in a number of initiatives to provide continuing support for evaluation, both globally and at the decentralized level. This includes the development of guidance, tools, approaches, and practical guidelines focusing in part on remote evaluation methods, rapid assessment techniques, quality assurance, and drawing out "lessons learned" regarding pandemic response.[11] The increased incidence of digital media since the start of the pandemic highlights the utility of strong national, regional, and global professional networks that have helped to keep the global evaluation community connected and current on shifts in approaches resulting from the crisis. Examples include the many Voluntary Organizations for Professional Evaluation across the globe, the Research for Development Impact Network (2020) webinar on M&E approaches during COVID-19, and the BetterEvaluation (Macfarlan, 2020) network. These avenues for information exchange and for developing emerging practice have highlighted several key features of

evaluation in the context of COVID-19, including: the higher reliance on remote data collection; increasing awareness of datasets that can be accessed digitally; reports of innovative means of stakeholder engagement and networks; and particularly the use of local evaluation specialists to conduct evaluation activities.

There is evidence in countries with a mature evaluation function of evaluation playing a "trusted advisor" role in supporting senior officials as they deal with the COVID-19 crisis.[12] This has required significant modifications in the practice of evaluation that can serve as useful "lessons" for the global evaluation community on potential approaches to ensure that the evaluation function is and remains relevant to decision-makers.

Evaluators at the country level have an opportunity to provide an important value-added role to national officials. Evaluation can help country officials gain a better understanding of their "most vulnerable," their needs, and possible solutions to improve their well-being. All evaluators will need to be nimble if they are to make inroads in providing value-added service to country-level decision-makers. Further adjustments are needed in the practice of evaluation, which has implications for the training and development of new and developing evaluators, assuming that evaluation and evaluators are seen by senior officials as a "need to have." Knowing how to balance an advisory role while also maintaining professional and ethical standards will become a critical element in the maturing of professional evaluators.

Furthermore, it is important to shift evaluation capacity development to the changes in MEL practice and encourage a transformed evaluation culture to suit the emerging context.

With international and domestic travel restrictions, there is a growing demand for national consultants, but there is also greater pressure on their availability. In one sense, this may accelerate the development of local, country-level evaluators. Additionally, the substitution of in-person learning events and conferences for webinars and other types of digital learning events has broadened their reach in terms of training and development opportunities in general and online discussions associated with the SDGs in particular. Far more emphasis is required on mentoring young and emerging national evaluators by senior evaluators. That said, there are also drawbacks, including the limitations put on experiential learning, which is so important to the development of new and emerging evaluators.

Table 5.3 presents a compilation of major themes observed from the experience to date of selected organizations and countries, inferred from the comments and analysis of various professionals and "experts" who represent a wide range of disciplines across the globe. Using the MEL SDG framework as the reporting guide, it suggests how the evaluation sector (nationally, regionally, and globally) needs to respond as the world emerges from the pandemic to best support the SDGs and country priorities in going forward.

Table 5.3 How the evaluation sector needs to respond to support the SDGs and country priorities.

Post-COVID Response of the Evaluation Sector
Vision and Commitment of Leadership
Successful examples of leadership's commitment to evaluation occur where evaluation provided a timely contribution to national decision-making, where senior country officials see the positive value in evaluation findings for decision-making, and the focus is on SDG outcomes rather than on specific SDG targets and indicators. For evaluation to remain relevant, evaluators will need to reinforce lessons shared by high-performing evaluation units during the pandemic: (i) being relentless about "client-centric" practice and (ii) ensuring that evaluation "has a seat" at the decision-making table. Evaluations need to generate learning in a manner that is appropriate to decision-making requirements. Lessons learned are: • Be proactive working with senior officials to raise awareness of the added value of evaluation; • position the evaluation function as part of the decision-making process; • generate guidance to support country evaluators and country officials on the various roles and uses of evaluation; • evaluators to position themselves with senior officials/decision-makers to understand the "big issues" and act as "trusted advisors;" • bring evaluation closer to decision-makers in a timely fashion to support rapid decision-making; • frame issues in the context of the SDG principles, even if indicators and targets have changed; and • re-think what SDG "success: might mean at the country level, for example, how digital access may offer new avenues of service delivery and employment, how food self-sufficiency in local areas is being diversified to provide nutritional balance, rather than the purchase of food.
Country SDG Strategy and Plan
The likelihood of SDG strategies and plans needing amendment due to COVID is high. Evaluation can make an important contribution, assisting in impact assessment on specific SDGs and principles; advising on re-setting of strategies or plans; and partnering with other stakeholders such as statisticians, auditors, and researchers. Evaluators can contribute evaluative thinking and tools to support collective decision-making in line with SDG principles. Strategic initiatives include: • Evaluability assessment to ensure that *theories of change* (TOC) reflect the changes that may have resulted from the pandemic and in response to the crisis; • on-going engagement in cyclical national planning, as well as shorter-term pandemic response initiatives; • closer examination of the SDG principles to better determine criteria for their assessment and gaining a common analytical understanding of them, for example, what does "leave no-one behind" mean? And how is it best measured?; • partnership and alignment with other analytical functions such as national statistics, audit, and policy research; • using TOC to assess environmental factors and other enablers that impact causal pathways to reflect the non-linearity and complexity of the real world; • evaluative thinking to support SDG implementation and examining multiple interventions, for example, through "nested" TOC; and • developing "Theories of Inequities" to facilitate the systematic assessment of the complexities associated with "leave no one behind."

(*Continued*)

Table 5.3 How the evaluation sector needs to respond to support the SDGs and country priorities. (*Continued*)

Post-COVID Response of the Evaluation Sector
National Monitoring and Evaluation System (NMES) Intensified efforts to strengthen NMES are needed for evaluation to systematically and proactively identify and address country needs and priorities post-pandemic. During the pandemic, important changes to evaluation practice occurred in the planning, roll-out, and management of evaluation. Evaluators have needed to be resourceful and nimble, moving beyond "traditional" practice. A challenge for organizations and evaluators may be in balancing the use of scarce evaluation resources between these various roles. Both practice standards and ethical guidelines for evaluation will need to be revisited to ensure an appropriate balance between the provision of consulting-type services using new, innovative, and rapid approaches, and quality standards of evaluation approaches. Additionally, there should be greater emphasis on communication across organizations and partnering with other analytical and research disciplines. Specific initiatives to be considered are: • Assist country officials with post-mortems on their COVID-19 response, assessing emergency preparedness and roll-out, and drawing "lessons" for the future. This could help address needs for policies, legislation, and responsibilities for future crises; • prepare evaluation frameworks to recognize core performance areas for COVID-19 response with comparable evaluation questions to allow aggregation of evidence across multiple exercises (i.e., to inform meta-studies and higher-level strategic analysis); • engage in more real-time results monitoring and analysis (i.e., process evaluations to provide timely and comprehensive feedback for adaptive management decisions); • recognize that "evidence" for decision-making is not singular; more coordination and collaborative approaches across other disciplines and experts is needed; • systematize information for ease of access and aggregation for higher-level analysis. Comparative analysis across countries, regions, and globally will help inform "good practices" for potential upscaling; • incorporate a focus on both broad context and changing dynamics of environmental considerations into evaluation analyses and reporting (effectively "zoom in" and "zoom out"); • balance the need to provide "good" evidence with delivering what is needed "just-in-time" for decision-making. Essential elements include: clarifying the scope (what critical questions need answering?), timing (when is information needed?), as well as developing a just-in-time approach for timely delivery; • place more focus on evaluation of processes to gain greater understanding of how progress is made, including the possibility of alternative approaches; • enhance communication across organizations, both within government and with stakeholders, to make best use of resources and technical support for addressing national priorities; • use new, innovative, and rapid approaches to evaluation (including forward-looking evaluative research) and new quality standards for evaluation approaches; and • phase planning, resourcing, and reporting of evaluations in an iterative manner to focus on critical aspects of implementation.

(*Continued*)

Table 5.3 How the evaluation sector needs to respond to support the SDGs and country priorities. (*Continued*)

Post-COVID Response of the Evaluation Sector
Enablers to facilitate "use" of evaluation by country stakeholders Short-term "lessons" from countries and organizations where evaluation has actively supported COVID response initiatives indicate the following "enablers" for evaluation use: a receptive senior management; good understanding by evaluators of the "big picture" and what is needed; and an experienced evaluation team that can deliver in a timely way. In countries and organizations where rapid feedback evaluation has successfully provided "just-in-time" support for COVID-19 response, demand is likely to grow rather than revert to "traditional" types of evaluation. Yet, other uses for evaluation ought not to be overlooked – for example, the in-depth learning role via developmental evaluation or an "accountability" role for evaluation. Scarce resources for evaluation mean that evaluators and evaluations need to demonstrate "value." Evaluators need to become more nimble, agile, and have greater foresight and adaptability than in the past. This also includes adjusting the content and approach to evaluation capacity building to broaden and update evaluation practice. This includes professional networks needing to recognize these emerging shifts in demand and practice requirements. This requires investment to ensure that sector changes occur broadly and rapidly. Examples of enablers to encourage the evaluation sector shift to the post-COVID-19 context include: • Evaluators need to clearly understand the different roles and possible uses of evaluation, placing greater operational focus on evaluation "use"; • establishing an appropriate balance between various possible uses of evaluation; • consideration of managing evaluation fieldwork without creating a "burden," real or perceived, on an organization, for example, through use of clustered evaluations; • greater focus on determining the point in fieldwork and analysis where information gathered is deemed "good enough," rather than relentless pursuit of perfection; • greater use of remote data collection and online communications for timely and contemporary data; • drawing from a broader and more diverse range of "lines of evidence," such as delivery of "rapid response briefings" to officials, including preliminary findings. This includes calls for developing new approaches to deliver "quick country assessments," "synthesis analysis," etc.; • partnering and coordinating across other disciplines and organizational units to deliver results within shorter timeframe, potentially increasing productivity, and broadening skill sets; • documenting and disseminating case studies where evaluators are pro-active and creative in using flexible approaches to maintain good evaluation outcomes; • updating evaluation training and training tools to incorporate new approaches to evaluation and evaluative research, including process mapping that examines trajectories in a broader context; • enhancing quality standards of evaluation approaches with new, innovative, and rapid approaches to evaluation, including forward-looking evaluative research; • evaluation networks providing mentorship and experiential learning to help evaluators – particularly new and emerging evaluators – to adapt quickly to new approaches and in bringing evaluative thinking to more complex issues;

(*Continued*)

Table 5.3 How the evaluation sector needs to respond to support the SDGs and country priorities. (*Continued*)

Post-COVID Response of the Evaluation Sector
• building a more intensive focus on evaluation "use" into formal training and experiential learning opportunities for new evaluators; • building increasingly sophisticated offerings of virtual training and development for evaluators; • strengthening the profile of evaluation and professional evaluators within the wider context of universities, research units, auditing, statistics, and other disciplines; and • encouraging collegial approaches to learning and mentoring across national, regional, and global evaluation communities, sharing on-the-ground lessons learned, and exploring how combined learning for new solutions can be found quickly. That said, it continues to be important to ensure that regional, cultural, and country-specific contexts get built into how evaluation is approached.

Summary of lessons from COVID-19 for evaluation and the SDGs

This paper has shown how the evaluation sector and MEL practice is already shifting to remain relevant in a post-COVID-19 world. For evaluation to support sustainable development and the vision of the 2030 Agenda, the authors have used the SDGs MEL framework to consider what, where, and how evaluation and practical MEL activities can contribute to re-establishing progress toward the intent of the SDGs.

The two key questions explored during this chapter highlight the following conclusions and emerging lessons.

What has COVID-19 meant for the use of evaluation in supporting the implementation, management, and reporting on the SDGs?

The COVID-19 pandemic has, to date, significantly altered the world of evaluation and its use in supporting the SDGs. The authors have pointed to the four key factors that influence the relationship between evaluation and the SDGs at the country level, noting that all four have been affected by the pandemic. Going forward post-pandemic, evaluators and the practice of evaluation need to adapt accordingly. A number of the changes realized during the period of the pandemic will likely endure. It will not be "business as usual" for the evaluation community but will require a shift that may lead to a "new normal."

Progress on most SDGs has been hindered by the pandemic since they have not been priority areas for attention. Yet the pandemic has created a much higher profile for the use of science and "evidence" in supporting decisions and actions of senior officials. This has not necessarily resulted in greater demand for systematic evaluation, but there is evidence in countries having a mature evaluation function of evaluation playing a "trusted advisor" role in supporting senior officials in dealing with the crisis. Evaluators now must find ways of providing more timely information to decision-makers, an area where evaluation has been criticized in the past.

The pandemic heightens the importance of monitoring, evaluating, and better understanding the SDG principles in terms of their transparency, equity, and universality, among other critical factors. In particular, the health, economic, and social impact of COVID-19 on the most vulnerable people in many countries has seen this group likely to increase. This raises the importance of analyzing the SDG principle of "leave no one behind." This impact is coupled with the likelihood that the current focus on SDG targets to be achieved by the year 2030 will have to shift. The global pandemic may have inadvertently raised both awareness and a sense of urgency among leaders of the need to deal with threats that are global in nature (such as SDG 13, 14, 15, and 17). In so doing, this could potentially serve to accelerate the commitment to these specific SDGs.

A key challenge for reinvigorating national evaluation capability-building efforts post-pandemic are the fiscal restraints, caused by the significant expenditures on COVID-19 response measures. It is important for the global evaluation community to champion the importance and use of evaluation to support SDG national planning and implementation, particularly in countries where evaluation capacity gaps currently exist. Initiatives such as the Global Evaluation Initiative of the World Bank and UNDP and efforts of UNEG, EVALSDGs and others will thus be important going forward.

Evaluation capacity building efforts need to involve much more than simply fine-tuning the curriculum for evaluators to include use of "non-traditional" approaches insofar as the planning, conduct, and reporting on evaluation are concerned. Digital learning events have broadened the reach of training and development opportunities and online discussions on evaluation topics help to build knowledge on important issues. However, limitations on experiential learning are a significant drawback for new and emerging evaluators. Awareness-raising for senior officials at the country level is critical, addressing where and how "evidence" and analysis generated by evaluation can be used for decision-making. A move to digital data collection mechanisms and other forms of social networking will bring fresh opportunities and challenges for the evaluation sector.

How can evaluation practitioners adapt in order to remain relevant to the SDGs and country-level needs?

The "lessons" from the pandemic show that evaluators need to be increasingly nimble if they are to make inroads in providing value-added service to country-level decision-makers – in particular, knowing how to balance an advisory role while maintaining professional and ethical standards. How evaluation practitioners adapt in the future to remain relevant to the SDGs and support country-level needs will vary somewhat across various countries. But there are also some commonalities on which the global evaluation community needs to reflect to develop support in going forward.

Evaluators need to engage with efforts to improve data systems, data availability (including more disaggregated data). Purposeful inclusion of the "leave no one behind" principle in evaluation requires new scrutiny to ensure that emerging vulnerabilities are recognized and addressed.

To remain relevant, evaluators need to increase the pace of evaluative initiatives in line with the needs of decision-makers; introduce new tools allowing for more rapid assessments. More emphasis needs to be placed on SDG principles and outcomes, but also understanding of shifting priorities, trade-offs, and new synergies. Evaluators need to be proactive in engaging with country officials to support post-COVID strategies targeting the revival of all sectors, using evaluation tools and evidence to support decision-making needed in weighing options regarding country transformation plans being considered as officials strive to "build back better" going forward. In communicating the benefits and importance of evaluation, using language that focuses on the practical benefits that evaluation brings is imperative. Evaluators need to use adaptive management practices such as context analysis and scenario planning while acknowledging shifting targets and reallocated resources.

National evaluation capacity building should be revitalized as a priority, incorporating on-line training, and training and orientation on the use of evaluation in the context of the SDGs. Support to national evaluators to lead country evaluations should incorporate experiential training of new and emerging evaluators, bringing in experienced evaluators as mentors.

Notes

1 The document "Transforming our World: The 2030 Agenda for Sustainable Development," or Agenda 2030, was endorsed unanimously by UN Member States in September 2015.
2 A wide variety of efforts to help fill national evaluation capacity gaps in support of SDG learning have been ongoing since (and prior to) Agenda 2030 launch by international agencies and global networks such as EvalPartners, EVALSDGs, UNEG, key UN agencies such as UNDP, etc.
3 The latest indicators and targets, with related metadata, are provided on the United Nations website (see, United Nations Statistics Division, n.d.).

4 Since 2015, 177 countries have generated VNRs, 118 for the first time, 48 have generated 2 VNRs and as of 2021, 11 countries will have generated 3 VNRs. This has included references to evaluation findings in VNRs; country-led evaluations supporting SDGs; national evaluation capacity development; etc.
5 Co-created by IIED, EVALSDGs, Ministry for Foreign Affairs of Finland, and UNICEF (see, D'Errico et al., 2020).
6 Several sources have served to focus on key imperatives for evaluation to support SDG implementation and reporting See, for example, EvalPartners (2015), Lahey (2016b, 2018a, 2018b), and D'Errico et al. (2020).
7 Given the recognized need in many countries for developing national capacity in evaluation and NMES, to enable the Agenda 2030 commitment to country-led evaluations in support of the SDGs, the state of maturity of the NMES and the need for evaluation capacity building are critical elements to assessing how COVID-19 may be impacting evaluation at the country level. See, Lahey (2015, 2016a) and UNEG (2015).
8 A summary of the event and associated links can be found online (see, United Nations, n.d.).
9 The UK Government, through the Office for National Statistic, has generated new, rapid data capture mechanisms. See, Office for National Statistics (2020).
10 See Bill No. 31, *Climate Change Act*, Parliament of the Republic of Fiji (2021).
11 Examples of UN agency initiatives include: UN Inspection and Evaluation Division (IED) of the Office of Internal Oversight Services (OIOS) that developed a COVID-19 response evaluation protocol that offers a conceptual framework for conducting the evaluation, with common questions, criteria, and performance indicators and measurements (see, Office of Internal Oversight Services, 2020b) and the Independent Evaluation Service (IES) of UN Women that has, among other things, developed a "pocket tool" that provides practical guidelines for gender-responsive evaluation management and data collection.
12 See Canadian Evaluation Society & Performance and Planning Exchange (2021a, 2021b).

References

Legal sources

Bill No. 31, *Climate Change Act*, Parliament of the Republic of Fiji. (2021). https://www.parliament.gov.fj/wp-content/uploads/2021/08/Bill-31-Climate-Change-Bill-2021.pdf

Secondary sources

Abidoye, B., Flex, J., Kapto, S., & Patterson, L. (2021). *Leaving no one behind: Impact of COVID-19 on the sustainable development goals (SDGs).* United Nations Development Programme & Frederick S. Pardee Center for International Futures. https://data.undp.org/wp-content/uploads/2021/04/Leaving-No-One-Behind-COVID-impact-on-the-SDGs-second-flagship-2.pdf

Anessi-Pessina, E., Barbera, C., Langella, C., Manes-Rossi, F., Sancino, A., Sicilia, M., & Steccolini, I. (2020). Reconsidering public budgeting after the COVID-19 outbreak: Key lessons and future challenges. *Journal of Public Budgeting, Accounting & Financial Management*, *32(*5), 957–965. https://doi.org/10.1108/JPBAFM-07-2020-0115

Canadian Evaluation Society and Performance and Planning Exchange. (2021a, March 4). *COVID, evaluation, performance measurement, and what comes next: Lessons from the pandemic – Session one* [Webinar]. Canadian Evaluation Society.

Canadian Evaluation Society and Performance and Planning Exchange. (2021b, March 17). *COVID, evaluation, performance measurement, and what comes next: Lessons from the pandemic – Session two* [Webinar]. Canadian Evaluation Society.

Committee on Statistics and Statistical Policy. (2020, May 26). *Impact of COVID-19 crisis on official statistics: Business continuity of official statistics*, OECD Doc. SDD/CSSP (2020)6. Organisation for Economic Co-operation and Development. https://unece.org/fileadmin/DAM/stats/documents/ece/ces/2020/SDD-CSSP_2020_6.en.pdf

D'Errico, S., Geoghegan, T., & Piergallini, I. (2020, January). *Evaluation to connect national priorities with the SDGs – A guide for evaluation commissioners and managers*. International Institute for Environment and Development. https://pubs.iied.org/sites/default/files/pdfs/migrate/17739IIED.pdf

EvalPartners. (2015, September 4). *Global evaluation agenda 2016–2020* [Draft]. https://evalpartners.org/sites/default/files/files/Global-Evaluation-Agenda-draft-4-September-2015.pdf

Florida, R. (2020, April 3). The geography of coronavirus. *Bloomberg*. https://www.citylab.com/equity/2020/04/coronavirus-spread-map-city-urban-density-suburbs-rural-data/609394/

G.A. Res. 70/1, 70th Sess., U.N. Doc. A/RES/70/1. (2015, September 25). https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_RES_70_1_E.pdf

Government of Costa Rica. (2018, April 25). *Report – Forum on data for sustainable development: "The roadmap for Costa Rica."* Global Partnership for Sustainable Development Data. https://www.data4sdgs.org/sites/default/files/services_files/Forum%20on%20Data%20for%20Sustainable%20Develompent%20-%20The%20Roadmap%20for%20Costa%20Rica_Report.pdf

Grossi, G., Ho, A., & Joyce, P. (2020). Budgetary responses to a global pandemic: International experiences and lessons for a sustainable future. *Journal of Public Budgeting, Accounting & Financial Management*, *32*(5), 737–744. https://doi.org/10.1108/JPBAFM-12-2020-189

Hale, T., Anania, J., de Mello, B. A., Angrist, N., Barnes, R., Boby, T., Cameron-Blake, E., Cavalieri, A., Di Folco, M., Edwards, B., Ellen, L., Elms, J., Furst, R., Green, K., Kira, B., Luciano, M., Majumdar, S., Nagesh, R., Petherick, A., … Wade, A. (2022, March). *Variation in government responses to COVID-19* (BGS-WP-2020/032). Blavatnik School of Government. https://www.bsg.ox.ac.uk/sites/default/files/2022-04/BSG-WP-2020-032-v13.pdf

Lahey, R. (2015, April). *Devising an appropriate strategy for capacity building of a national monitoring and evaluation system: Lessons from selected African countries*. World Bank. http://hdl.handle.net/10986/22079

Lahey, R. (2016a). Why developing monitoring and evaluation capacity is critical to understanding and addressing issues of poverty and inequality. In R. C. Rist, F.P. Martin, & A.M. Fernandez (Eds.), *Poverty, inequality, and evaluation: Changing perspectives* (pp. 209–235). World Bank. http://hdl.handle.net/10986/22769

Lahey, R. (2016b, December). *Applying evaluative thinking towards an effective ILO contribution to the implementation of the SDGs*. International Labour Office. https://www.ilo.org/wcmsp5/groups/public/—ed_mas/—eval/documents/publication/wcms_537378.pdf

Lahey, R. (2018a, February). *Diagnostic instrument to assess the evaluability of the DWCP in the context of the SDGs*. International Labour Office. https://www.ilo.org/wcmsp5/groups/public/—ed_mas/—eval/documents/publication/wcms_625970.pdf

Lahey, R. (2018b, September 11). *Using evaluative thinking to support effective implementation of the SDGs: It's not business as usual* [Webinar]. EVALSDGs. https://evalsdgs.org/2018/09/11/webinar-using-evaluative-thinking-to-support-the-effective-implementation-of-the-sdgs-its-not-business-as-usual-for-evaluators/

Macfarlan, A. (2020, April 21). *Adapting evaluation in the time of COVID-19 – Part 1: MANAGE*. BetterEvaluation. https://www.betterevaluation.org/en/blog/adapting-evaluation-time-covid-19-part-1-manage

Matambo, O. K., Botswana, & United Nations in Botswana. (2018). *Botswana SDGs roadmap: A guideline to implementation of sustainable development goals, 2017–2023*. Ministry of Finance and Economic Development in collaboration with the United Nations Botswana

Multilateral Organisation Performance Assessment Network. (2021, October). *Lessons in multilateral effectiveness – Pulling together: The multilateral response to climate change*. https://www.mopanonline.org/analysis/items/MOPAN_MLE_CC_Brief_web.pdf

Office for National Statistics. (2020, March 27). *Ensuring the best possible information during COVID-19 through safe data collection*. https://www.ons.gov.uk/news/statementsandletters/ensuringthebestpossibleinformationduringcovid19throughsafedatacollection

Office of Internal Oversight Services. (2020b, October). *COVID-19 response evaluation protocol*. United Nations Evaluation Group. http://www.unevaluation.org/document/download/3687

Organisation for Economic Co-operation and Development & United Nations Development Programme. (2019, November 1). *Making development co-operation more effective: 2019 progress report*. https://doi.org/10.1787/26f2638f-en

Research for Development Impact. (2020). *M&E approaches during COVID-19*. https://rdinetwork.org.au/resources/me-approaches-during-covid-19/

Sustainable Development Solutions Network Secretariat. (2014, November 20). *Framing sustainable development goals, targets, and indicators*. https://irp-cdn.multiscreensite.com/be6d1d56/files/uploaded/141120-Framing-Goals-Targets-and-Indicators.pdf

United Nations. (n.d.). *High-level political forum 2021 under the auspices of ECOSOC*. https://sustainabledevelopment.un.org/hlpf/2021

United Nations. (2020a, April). *A UN framework for the immediate socio-economic response to COVID-19*. United Nations Sustainable Development Group. https://unsdg.un.org/sites/default/files/2020-04/UN-framework-for-the-immediate-socio-economic-response-to-COVID-19.pdf

United Nations. (2020b, April 9). *Policy brief: The impact of COVID-19 on women*. United Nations Women. https://asiapacific.unwomen.org/en/digital-library/publications/2020/04/policy-brief-the-impact-of-covid-19-on-women

United Nations Department of Economic and Social Affairs. (2019). *2019 voluntary national reviews synthesis report*. Sustainable Development Knowledge Platform. https://sustainabledevelopment.un.org/content/documents/252302019_VNR_Synthesis_Report_DESA.pdf

United Nations Department of Economic and Social Affairs. (2021). *2021 Voluntary national reviews synthesis report*. Sustainable Development Knowledge Platform. https://sustainabledevelopment.un.org/content/documents/294382021_VNR_Synthesis_Report.pdf

United Nations Development Programme & World Bank. (2016, November 9). *Transitioning from the MDGs to the SDGs*. United Nations Development Programme. https://www.undp.org/sites/g/files/zskgke326/files/publications/Transitioning%20from%20the%20MDGs%20to%20the%20SDGs.pdf

United Nations Evaluation Group. (2015, September). *National evaluation capacity development: Practical tips on how to strengthen national evaluation systems*. United Nations Sustainable Development Group. https://unsdg.un.org/sites/default/files/UNEG-National_Evaluation_Capacity_Development-Practical-Tips-ENGLISH.pdf

United Nations Statistics Division. (n.d.). *SDG indicators – Metadata repository*. United Nations. https://unstats.un.org/sdgs/metadata/

6 Implications for Evaluation, What We Learn from the UN and Country COVID-19 Response Plans, and Reflecting on Future Scenarios

Indran A. Naidoo

Introduction and background

Historically, evaluation has evolved in a context of supply and demand, with the form of evaluation and the priority accorded to it based on the requirements given. In the development context, international funding and donor interests drove the profession, with an emphasis on accountability. The forms of evaluation that measured results to inform policy, funding decisions, or both, drove the development of the practice. Such an approach situated evaluation as a form of audit for the recipients, who often perceived the process as critical and punitive rather than constructive and beneficial. This reluctance to engage meaningfully accounts in part for the slow uptake of evaluation by most countries globally, as the profession was initially informed also by the Eurocentric evaluation literature that failed to explain its value accountability. Misperception and misinformation persist today in the way international bodies use evaluation, effectively serving as an additional assurance or fidelity function to governance bodies (Schwandt, 2019).

The Paris Declaration on Aid Effectiveness[1] was an important shift, as it emphasized self-determination and sovereignty, and began a process where evaluation became less donor-driven and more country-driven. Through the efforts of the United Nations Development Programme (UNDP), its convening power has been used to stress how evaluation is linked to concrete outputs and outcomes. As a result, the receptivity to evaluation has improved and the emphasis on accountability has decreased (Wilton Park Dialogue, 2018). This, in turn, has increased the interest of governments in the potential of evaluation, given that its framing had shifted away from funding conditionality. Geo-political shifts, together with a recognition that evaluation can benefit countries by helping them improve their development effectiveness, have also contributed to evaluation being embraced over time. The time when a discrete evaluation report from an independent or credible evaluation office should be viewed as a definitive form of uncontested judgement has passed.

DOI: 10.4324/9781003376316-7

COVID-19 has brought another major shift to evaluation, as Jan Eric Furubo argues in his chapter in this book. The evaluative conversations in this new construct will be informed by multiple information providers and actors and occur across different platforms and modalities (Rist & Stame, 2006). The exclusivity held by evaluators, irrespective of reputation or credibility, will change as their voices will be but one of many informing evaluation conversations. This chapter examines some of the changes that have occurred and that currently influence evaluation, for the perspective of its reconfiguration in a post-COVID era, its engagement with other research providers who have entered the evaluative space, and a reflection on what this means for evaluation functions and professionals in the future.

Evaluation emphasis changing from accountability to supporting SDG attainment

Over the past two decades, and especially in the last decade, evaluation developed in two fundamental ways. First, it has moved from being something which international players and donors insisted on, to being driven at a country-level and by civil society actors. Second, evaluation's value proposition has transitioned from being merely accountability-oriented to supporting policy formulation and promoting learning, especially towards the 2030 Agenda for Sustainable Development ("Agenda 2030"). All countries are now actively engaged in the global evaluation-development space and provide multiple models of leadership by introducing new philosophies and forms of evaluation. As a result, there is a greater sense of ownership of evaluation at the country-level.

The UN's convening power has undoubtedly advanced the role of evaluation in promoting progress towards the attainment of the Sustainable Development Goals (SDGs), which has emphasized national leadership and ownership of evaluation and has sought to improve the quality of national evaluation systems. The National Evaluation Capacity (NEC), organized by the Independent Evaluation Office (IEO) of the UNDP, found that the majority of the 160 participating governments had improved their use of evaluation to advance towards Agenda 2030 (Naidoo, 2020b, pp. ix–xi) cumulatively over the last decade.

The UN engages formally with governments through its convening power, with the NEC series being a central platform. In these events, it has engaged most evaluation networks around key topics such as the SDGs, evaluation criteria, and methodology, and has provided training to participants. Networks included the Evaluation Cooperation Group (ECG),[2] the United Nations Evaluation Group (UNEG),[3] and the EvalNet.[4] The meetings have introduced contemporary thinking on evaluation matters and have helped build capacity for government participants.

Government participants used the NEC platform to share their experiences in using evaluation, to report on and advance the attainment of the SDGs (Naidoo, 2020b, pp. ix–xi). The discussions indicated the achievement of maturity over time concerning earlier work and themes. These themes began from evaluation foundation-building and setting evaluation policies to more substantive discussions on the role of evaluation and how it was used in a more grounded manner to advance development agendas, and measure and report progress towards the SDGs (Naidoo, 2010, pp. 303–320). Evaluation moved towards a focus on providing practical solutions (Schwandt, 2005, pp. 95–105). The result was increased capacity to support the attainment of Agenda 2030 (van den Berg et al., 2017).

The progress of the NEC series over the past decade reflected these changes, moving evaluation in a direction that was action-oriented and pragmatic, whilst supporting future planning. First, it explored themes such as evaluation as a public good (2009, Morocco), followed by evaluation and public policy (2011, South Africa), and then advanced discussions on the implications for principles of evaluation (2013, Brazil). It progressed from a focus on evaluation as a tool to demonstrating how good evaluation helped improve people's lives (2015, Thailand). The discussion then incorporated how UN efforts to promote development through the SDGs was a way of measuring progress (2017, Turkey). The last event took stock of what was achieved in terms of evaluating SDG-evaluation attainment (2019, Egypt).

Baseline pre-COVID and changes in the COVID-19 era

In global discussions, the NEC has served as a forum for highlighting fresh ideas. In 2019, for example, it convened an event on good practice standards (Naidoo, 2020a, pp. 63–69) and a case study from the IEO helped countries reflect on their own evaluation evolution (UNDP, 2020). The forum identified the following building blocks of good practice, although changes are likely to occur in the new era, due to the reprioritization that will be inevitable as we move forward:

- Evaluation policy, including governance and funding;
- independence, objectivity, and the SDGs;
- quality assurance of evaluation (UNDP, 2019);
- collaboration between evaluation and audit (Naidoo & Soares, 2020);
- addressing substantive needs and demands;
- evaluation scope; and
- communicating evaluation (Universität Bern, 2018).

The extent to which governments have embraced these aspects of good evaluation practice vary, reflecting the differences in evaluation systems across and within countries. Although there has been progress over time, key aspects such as the structural independence of the evaluation units

were not always realized. Other factors also prevent optimal performance for evaluation functions, such as the lack of clear policies, reporting lines, and budgets. The evaluation models employed also varied, with some governments outsourcing to draw in evaluation capacity, whilst others relied on their own capacities. The IEO case studies are important in that they illustrate a high degree of variability in evaluation capacity, which did not place evaluation in a strong position during the COVID-19 crisis to be able to support needs with maximum impact.

Country-level evaluation mechanisms may be challenged due to COVID-19

Over the last two decades, significant effort and resources have been invested in developing evaluation principles, policies, and practices at the national level. The growth of evaluation has resonated well with progressive ideals of advancing and supporting democracy, transparency, and accountability. There has been significant uptake of the practice within countries and their governments following recognition of the potential value of evaluation in decision-making and performance improvement. The evaluation sector has responded by supporting the development of evaluation architecture within and across countries, with multiple complementary global efforts linking evaluation to the attainment of normative and development goals. This has helped infuse an evaluation discourse into the planning processes of governments and raise awareness about the importance of being able to measure and respond to results, whether derived from political or administrative commitments.

Substantial progress has been made to embed evaluation across all sectors. In particular, it has helped to build the practice through dedicated occupational categories for evaluation-related activities in governments. There have been advancements towards professionalization accompanying an expansion of evaluation networks and associations. There is an extensive dedicated literature on the subject, illustrated by the number of books, journal articles, and diversity of experience demonstrated in the multiple mediums. Along with the political, civic, and administrative systems that advance the practice, there are greater efforts to systematically build and use evaluation capacity. The demand for accountability also comes from citizens who wish to see credible reports of results (Naidoo, 2004, pp. 8–11). Numerous evaluation networks and associations reflect the priorities of different evaluation constituencies, including consultants and evaluation professionals, commissioners, government users, academia, and civil society. All share the common ideal that evaluation seeks to make a difference by improving performance. As part of oversight, and together with audit, evaluation has been driven by criteria that aim to optimize the use of resources, promote efficiency and effectiveness, measure relevance and sustainability, and create value (Naidoo, 2020c, pp. 177–189).

Changing realities may alter how evaluation is conducted

Some features of the pre-pandemic evaluation architecture do not align well with the requirements for information now needed by countries. The shifts during the COVID-19 crisis occurred because traditional evaluations were seen as increasingly outdated along with the actors who understood the evaluations. The shifts come from department- or agency-specific approaches towards *holistic country approaches*, using multi-sectoral integrated approaches within sector-specific decision-making, based on multiple streams of information (compared to traditional evaluative information, which is generally discrete and based on singular reports). Whilst the discussions during the pandemic suggested joint evaluations and information-sharing between departments and agencies (UNEG, 2020), the linkages to other departments or agencies, or evaluating "as one," remains largely non-existent. This more siloed approach has limited the relevance of reports in all-of-government or all-of-society approaches, both of which are key principles stated in the UN COVID-19 socio-economic response (United Nations, 2020) in its efforts to "Build Back Better" (United Nations, 2020). The pooling of development resources to support recovery efforts assumes that evaluation capacities and resources should be blended. In practice this has not happened, as evaluation functions continue to operate in a siloed fashion, serving the more focused needs of various agencies and their governing councils.

These examples illustrate the shortcomings of an evaluation architecture that, despite its evolution in recent years, continues to be linear and simplistic, and assumes a high degree of predictability and stability. It also assumes regular funding flows premised on predictable budgets (including taxes, remittances, and Official Development Assistance), predictable growth rates based on historic trends, and overall optimism. Today, this predictability is lost, and the operational environment of evaluators has altered, as evaluators are now competing with new actors at the country level. Institutions with strong academic and research capacities have gained considerable traction in providing oversight services.

These academic and research institutions possess strong multidisciplinary networks and can produce comprehensive work of an evaluative nature. They may potentially challenge smaller evaluation units that do not possess such capacities or networks. Evaluation curricula have become more developed and strong support has been provided to build the skills of people who train as evaluators. Research institutions can draw on and harness the latest technology to access large databases needed for appropriate assessments of the scale and magnitude of development questions at the country level. Major research institutions are also able to draw on real-time and disaggregated data to conduct scenario planning. Government users need such information that can be provided at low or no cost in a rapid manner, and that can be focused on real socio-economic development

challenges (Voccia, 2021). In addition, technology has the capacity to replace the need for physical interviews and other ground-truthing, which may replace a key element that evaluators had not only used to support their professional role and legitimacy but also for verification and deeper understanding. This further reduces the opportunity cost of using national-level evaluation capacities over established evaluation outfits.

Findings from a review of the SERPs as it relates to evaluation

This section of the chapter highlights some findings from the UN review of the socio-economic response plans (SERPs) and its the implications for evaluation in the future. These new review and planning processes, installed by the UN and government compacts across 140 countries, claim to be collaborative and work horizontally, and emphasize issues that should promote recovery like human rights and inequality. The UN review used a rubric to assess comprehensiveness and the extent to which the plans were data informed. It also examined economic performance and impact on population groups and focused on humanitarian crises, the environment, economic dependencies, and the impact of value chain disruptions.

What the plans seek to achieve and their focus

The SERPs are joint government-UN documents, agreed by both parties. They seek to be comprehensive and emphasize joint responsibility for results. The policy and guidance documents intend to provide an empirical and logical basis for designing new development pathways. They seek to instil global normative values and priorities into the national sphere. The instrument claims to focus on response and planning efforts, and to be people-centred, whilst allowing countries to work out implementation modalities.

Each of the five pillars of intervention contains baseline data from which interventions may be monitored, and theoretical scenarios based on the severity of the crisis. The data to support the interventions is drawn from existing and planned studies to ensure that interventions are effective.

Discrete evaluation reports have little value in a collaborative context. Furthermore, evaluation entities do not have the ability to work within and address the comprehensive nature of the UN COVID-19 SERPs. This envisages a degree of joint leadership and funding for securing data, as well as developing a common understanding of what potential changes the crisis will require. The joint approach must also include an understanding and response to the humanitarian-environmental nexus and track deviations from SDG targets which could derail progress. Developing policy options to address vulnerabilities and inequalities with the intention to address structural inequality is also important.

The existing country-level evaluation approaches, even from the international evaluation offices, cannot provide the comprehensive approach required to deliver meaningful policy options. Evaluation has been largely absent in this development at the country level.

There are indications that the traditional evaluation ecosystem shall change

Whilst evaluation gained prominence for its potential role in supporting the attainment of the SDGs and progress towards meeting Agenda 2030, the COVID-19 crisis introduced new realities. Aspects of evaluation that were necessary for supporting the SDGs, such as a strong evaluation architecture, clear deliverables of products to inform SDG progress, and the resources to deliver these activities, no longer fit into a development planning paradigm. The new development priorities and more comprehensive ways of working triggered by the COVID-19 crisis rendered evaluation, in its current form, less effective. This is because evaluation is not configured to be agile and responsive and has generally worked by supporting discrete mandates or features that do not help in this new context. Nonetheless, strengthening local capacity for measuring progress on the SDGs remains important given the interlinked relationship between the SDGs and national development goals in many countries (UNDP, 2020).

Government responses as gleaned from the SERPs

The choice of shutting down the economy to save lives has been a source of tension and was hard to justify in the absence of economic measures to support the loss of incomes. Governments' ability to manage these conflicting goals was regularly challenged. It has brought to the surface questions of how well governments are able to address the humanitarian crisis. Whilst most claim that their response was evidence-based, the absence of sufficiently transparent monitoring and reporting systems means that health-protection plans are largely aspirational. There has been little publicly accessible evidence of progress based on what the plans have set out. The SERPs mention oversight committees, with collaboration and joint responsibility for results and reporting. Many government plans have included expanded membership to encompass academia, non-governmental organizations, and the private sector. Most SERPs include, as a minimum, the UN and government leadership at the country level, working to execute the plans jointly.

The extent to which governments respond to these factors, as described in the SERPs, will only be known through an independent monitoring and evaluation system with both national and international credibility to answer questions about the effectiveness of the COVID-19 measures. Examining whether scarce resources are targeted to the vulnerable and the

poor, whether new forms of economic activity have succeeded in reducing reliance, and whether "Building Back Better" does in fact happen, cannot be known based on inward-looking monitoring and evaluation systems. These are difficult political questions, especially as systems are unlikely to develop to the standard required given the political sensitivities associated with reporting transparently on government effectiveness during COVID-19.

The SERPs reports show that governments are already using the research capacity of universities to assist in the planning process, as they have access to other forms of data. However, one of the major challenges is the lack of data on key sectors of the economy, especially those most impacted by the pandemic, such as the informal sector, which makes up 60 per cent of the global workforce (an average figure, which is likely significantly higher in many lower-income countries) (United Nations, 2020, p. 17). The lack of disaggregated data by gender and other vulnerability markers has made it difficult to identify equitable solutions. These groups tend to be outside formal structures and face other levels of danger and vulnerability, such as discrimination and marginalization.

It cannot be assumed that information technology and internet connectivity will solve these data problems. There remains the problem of electricity, equipment, and connectivity costs; the digital divide is a hindrance in most countries. The COVID-19 crisis may have made the digital divide and many of its digital requirements worse. Working remotely is not an option for the informal and service sector and a nuanced approach to evaluation using virtual techniques would be required to include the most marginalized.

What the SERPs suggest about an emerging evaluation architecture

The SERPs indicate that the current oversight architecture is inadequate, as it has traditionally operated in a predictable rather than dynamic, crisis context. There is limited capacity for working across oversight structures or understanding that oversight can be a comprehensive process which is collaborative rather than mandate driven. These are multi-year national development plans, most of which contain references to monitoring and evaluation as a means to periodically assess progress. Whilst the plans are national in nature, they reflect a siloed approach of individual ministries, many of which do not collaborate.

As for UN interventions, they are evaluated by UN agency-specific offices, and results do not feed into a broader evaluation discussion. There has been limited UN agency collaboration and few efforts to change this through a new coordination system; sustained results are yet to be demonstrated.

The SERPs suggest that the shift taking place has an emphasis on more actors reporting on progress. These national actors acquire a crucial role

in country-driven evaluations and the evaluation practice suggested by the SERPs requires them to understand the scale of the changes promoted by interventions and to respond accordingly. The plans also emphasize the complexity of the current reality and the importance of taking it into account for policy development. Traditional evaluation practice, with its linear orientation and mandate-specific focus, is not agile enough to produce the insights necessary for the new context.

The existing evaluation architecture may no longer be relevant to shifting priorities and may be uncoordinated with new ways of working. If the SERPs are indicative of the future, there will be less emphasis on information from agencies and a greater focus on "conversation-driven," collaborative and engaging work. The evaluation architecture will be less definitive and more focused on the future compared to the past.

All country development plans have been reframed to ensure their relevance to recovery efforts. The international community, too, has had to reassess how it measures its intervention success. The previous plans were purely focused on a development pathway for Agenda 2030 and the attainment of the SDGs (Naidoo & Soares, 2017, pp. 51–63). Now, the SDGs remain important but take on a new emphasis; joint guidance by the OECD and UNDP, for example, has positioned the pandemic as an opportunity to "spark a new wave of innovation and ambition" relating to Agenda 2030 and the SDGs (Independent Evaluation Office/United Nations Development Programme & Organisation for Economic Co-operation and Development/Development Assistance Committee, 2020).

The response by evaluators during the crisis with evaluation

During the pandemic, pockets of evaluative activity have focused on alternative methodologies to support old practices rather than on recognizing the magnitude of the crisis and its future implications. The oversight contexts have changed alongside new demand and supply sources for evaluation. As previously noted, the academic and research sector has stepped into the space left by the inertia of evaluation leadership; with the limited visibility of evaluators on the frontline of an historic global crisis have, it is unclear what the future holds for monitoring and evaluation (The Wits School of Governance, 2020).

The review of over 80 SERPs (Naidoo, 2022) shows that the traditional oversight architecture no longer functioned as usual during the pandemic, but there has been significant growth in the offering from research institutions to governments to manage the response. There has been little contribution of evaluation expertise from the international evaluation networks to support recovery efforts, either through their agencies or collectively.

The fracturing of the stable environment that was conducive to an effective evaluation architecture and the inability of the international evaluation community to either respond creatively or recreate itself means this space is effectively lost.

Evaluators should be able to adapt to the new development and evaluation discourses

In the future, evaluators will need to expand their frame of reference and to understand and be able to work with complexity. Evaluators have tended to be reductionist in their approaches, simplifying complex issues in a manner that is hard to justify, using outdated methodologies, and generally being unable to evaluate beyond their agency mandates.

Evaluative skills require conceptualization at the global, regional, and country levels. They also require an understanding of the scale and interaction amongst various levels, being able to frame assessments in the context of political and developmental issues, and the ability to construct policy options. More specifically, and based on the SERPs review, evaluators must be able to frame the COVID-19 crisis globally against the backdrop of previous development trajectories and inherited vulnerabilities.

Each of the content areas includes a set of interconnections, which are complex and part of a fast-changing dynamic which is inherently political and influenced by geo-political factors. In addition to the factors already mentioned, there is the digital divide, the role of the diaspora in the context of population movements and migration, and changes to the operation of financial development institutions, including what the COVID-19 crisis means for debt and other obligations. Issues of food security, triggered by the closing of markets and disruption in production contribute to the complexity. Projecting ahead, major additional research capacity and streams of information will be needed, the most obvious being shifts away from singular agency or departmental evaluative reports for discrete audiences towards reports from established institutions. These should address the complexity and the nature of the comprehensive information required. This development emerges relatively well from think tanks and research institutions, which are also strong in providing multi-disciplinary perspectives.

Evaluators need to understand and work with scale

At the broadest level, the COVID-19 crisis amplifies existing inequalities and levels of differentiation. A response that is generic and presents an aggregated reaction will mask these inequalities and disproportionate impacts. To address this, however, would require sophisticated data and analysis, something not generally present in individual evaluation units. The emphasis on singular interventions, which is a feature of

agency-specific evaluation units (at the government level and those within the international evaluation networks) translates to their inability to deal with differentiated impacts across scale.

The SERPS review shows that interventions need to reach beyond urban communities to focus on any disproportionate impacts between and within peri-urban and rural populations. Many of the targeted populations do not benefit from public-sector infrastructure, making it difficult for them to access services. A key factor identified in the review is the digital divide, with the lack of electricity, funds, computers, and networks preventing remote education; tele-medicine; and other digital service provision. Therefore, the ability to evaluate across different geographic levels and scales means that the deeper levels of socio-economic differentiation are glossed over, in part due to the use of averages. Census data is not comprehensive enough to allow for pro-poor targeting. This means that most of the policy options and pro-poor policies will lack the benefit of solid insights. The reviews also point to increased discrimination based on gender and other grounds. However, in the absence of solid monitoring data, the real impact is unknown.

The implicit capital or headquarter bias in government operations is mirrored by urban and official data bias. Data tends to be aggregated and all the SERPs demonstrate a deficit in disaggregated data, even at the level of national census data. If information fails to highlight variations within the population, the policy responses will fail to address social and economic differentiation adequately. The policy options presented as interim responses in the SERPs have already shown a bias towards the aggregated data. Framing the response according to scale means moving beyond the comfort zones of the capitals and governments and generating information on historically marginalized. This is often not possible.

Critical content areas that require specialized knowledge for monitoring and evaluation

The COVID-19 crisis has multiplied the number of development challenges. Until the crisis, the SDGs served as a comprehensive set of common indicators to measure progress. The magnitude of the crisis resulted in much deeper changes that fundamentally affected established systems, including evaluation. It has been observed that the crisis paused regular activities, such as the established practice of reporting government progress against set plans, especially in countries with some form of democracy. The de-prioritization of this form of accountability, given the crisis context, has created questions which need to be asked if there is to be a reestablishment and reorientation of evaluation practice as an accountability measure.

The reorientation of evaluation to work across sectors and agencies, producing high-quality real-time evaluative information for immediate recovery

The evaluation sector was developed in environments which had a degree of predictability; the planning processes of governments assisted in fostering that stability, with a clearly established sets of users for results reporting. The demand and supply for this type of evaluative work, however, has taken place in silos, as mentioned above, and there has been very little horizontal collaboration in oversight. During the crisis, the established systems were interrupted. Resources were pooled and reprioritized, and reporting on results was no longer the sole preserve of any agency; rather, it became a joint collaborative reporting effort.

The SERPs mention evaluation, but with limited details, and many of the reviews come from non-traditional evaluation sectors such as research and academic think-tanks. They have been able to deliver at the speed and scale required. Good examples include the National Council for the Evaluation of Social Development Policy in Mexico and the National Institution for Transforming India (NITI) Aayog in India. They marshal national level evaluation capacities from academia as well as the public sector. Such institutions are also best placed to provide institutional legitimacy if required when it comes to making evaluative judgements. Whether the evaluation units of government or international agencies can contribute to this new space shall become evident over time.

Evaluation units tended to be small compared to these other entities, and the absence of significant responses during the pandemic likely indicates a lack of preparedness or an inability to retool evaluation to meet new needs and demands. Evaluation is not as familiar with big data and geo-spatial analysis tools and this gap has been evident in the reports produced by the academic and research sector. This work may not meet the standards of evaluation in terms of independence, but it has met research standards and has been able to deliver results.

The COVID-19 crisis has thus uncovered many weaknesses in the evaluation system. The focus on discrete interventions has no value in an all-of-government or all-of-society approach. This requires major ideological and behavioural changes from evaluation. Historically, evaluations have remained disengaged from policy and operational interventions; the new context requires more engagement.

Evaluators need to work at multiple levels and be able to unpack disaggregated data

There are three principal and interrelated challenges in implementing the response to COVID-19: equity, public sector capacity and data availability. In terms of *equity*, as the SDGs clearly articulate, it is of utmost

importance to reach all groups of people, especially those who are most vulnerable. In a post-COVID world, it becomes even more crucial given that the crisis has hit the informal sectors and vulnerable population the hardest. Implementing equity-focused programming requires robust *public sector capacity*. The review of the SERPs has highlighted weaknesses within the public sector infrastructure, which tends to be urban-biased. Thus, weak public sector capacity has an adverse impact on equity. COVID-19 has revealed the limitations of centrally driven, top-down approaches to programming and evaluation. The final factor, which is important for programming and evaluation in a post-pandemic world, is the availability of *reliable data*. This factor, linked to research capacity within countries, openness to alternative data sources and views, and media freedom is critical for evaluation.

What is required of evaluators in the new context?

At the country level, there has been at least some collaboration amongst the various agencies. A new UN system to improve coordination has shown to be effective, as reflected in the SERPs, which emphasize the pooling of resources, at the very least from the UN, towards a joint UN approach (Freeman et al., 2022). As the substance of the SERPs shows, the following attributes are required for any oversight and support function:

- Agility and the ability to work across mandates collaboratively and evaluate as one. The various pre-crisis efforts to bring about evaluation coordination to mirror the reform efforts seeking to get the agencies to work seamlessly were unsuccessful.
- Participation in the efforts for actual collaboration, reprioritization, and commitment to joint budgeting by agencies and departments, evidenced by the SERPs.
- Possession of specialized content knowledge and understanding required in the light of the new development context, before moving into developing monitoring and evaluation systems. There has not yet been an audit of the skills of evaluators against the new content focus areas at the country level. It is evident that the institutional capacity to provide the content proficiency and work to scale is more present in the academic and research fields than in the evaluation sector.
- Capacity for co-creation of knowledge and working in a holistic, all-of-society approach, which is generally not within evaluators' experience.

It was assumed that engagement would compromise the ability to provide objective judgement. The other shift in understanding is that past trends no longer offer any reasonable basis on which to offer propositions,

whether recommendations or insights. The magnitude of the crisis has been such that the focus is on immediate recovery efforts, working in challenging and under-resourced contexts with little time available to await long reports.

Conclusion

The COVID-19 crisis has posed challenges and opportunities for evaluation. This chapter argues that the main response to these challenges has been superficial and methodological, employing stopgap measures to mitigate an inability during the pandemic to gather real-time, credible information that assists in decision-making. The use of tools like geographic information systems and remote sensing is part of a technological addendum to evaluation but cannot replace the need for what remains a strategic and analytical function (García & Kotturi, 2019). There have been shifts in the governance environment which has affected the evaluation architecture, which has been relatively secure for supporting fidelity evaluation. The need for classic accountability evaluation shall change as funding alters alongside geo-political shifts that call for more self-determination of evaluation. The inability of the evaluation community to adapt its value proposition and enter the new development space however is concerning and may affect its further relevance.

Notes

1 The Paris Declaration was endorsed at the Second High Level Forum on Aid Effectiveness in 2005. It is a practical, action-oriented roadmap to improve the quality of aid and its impact on development (Organisation for Economic Co-operation and Development, 2005).
2 The Evaluation Cooperation Group is the professional network of the World Bank and regional banks.
3 The United Nations Development Group (UNEG) is the professional networks of the evaluation and oversight offices of the United Nations.
4 The DAC Network on Development Evaluation (EvalNet) is the evaluation network of the bilateral agencies and is led by the OECD Development Assistance Committee (DAC).

References

Freeman, T., Esser, A. L., & Vela, P. (2022). *Interim report: System-wide evaluation of the UNDS response to COVID-19.* United Nations. https://unsdg.un.org/sites/default/files/2022-03/Interim%20Report-%20SWE%20of%20the%20UNDS%20Response%20to%20COVID-19_Final.pdf

García, O. A., & Kotturi, P. (Eds.) (2019). *Information and communication technologies for development evaluation*. Routledge.

Independent Evaluation Office/United Nations Development Programme & Organisation for Economic Co-operation and Development/Development

Assistance Committee. (2020, April). *Good practices during COVID-19*. United Nations Development Programme. http://web.undp.org/evaluation/guideline/documents/covid19/IEOOECD_DAC_Joint-Guidance_COVID19.pdf

Naidoo, I. (2004, November/December). *The emergence and importance of monitoring and evaluation in the public service*. Public Service Commission News. http://www.psc.gov.za/documents/pubs/newsletter/2004/psc_news_book_nov_2004.pdf

Naidoo, I. (2010). Monitoring and evaluation in South Africa: Many purposes, multiple streams. In M. Segone (Ed.). *From policies to results: Developing capacities for country monitoring and evaluation systems* (pp. 303–322). UNICEF. http://www.theoryofchange.org/wp-content/uploads/toco_library/pdf/2010_-_Segone_-_From_Policy_To_Results-UNICEF.pdf

Naidoo, I. (2020a, October). Architecture for evaluation effectiveness. In Independent Evaluation Office of UNDP. *Proceedings from the 2019 national evaluation capacities conference*. United Nations Development Programme. http://web.undp.org/evaluation/documents/NEC/2019/NEC2019_proceedings.pdf

Naidoo, I. (2020b, October). Foreword. In Independent Evaluation Office of UNDP. *Proceedings from the 2019 national evaluation capacities conference*. United Nations Development Programme. http://web.undp.org/evaluation/documents/NEC/2019/NEC2019_proceedings.pdf

Naidoo, I. (2020c). Audit and evaluation: Working collaboratively to support accountability. *Evaluation*, *26*(2), 177–189. https://doi.org/10.1177/1356389019889079

Naidoo, I. (2022). *Evaluation for transformational change*. Unpublished manuscript.

Naidoo, I., & Soares, A. (2017). Incorporating the sustainable development goals in national evaluation capacity development. In R.D. van den Berg, I. Naidoo, & S. D. Tamondong (Eds.). *Evaluation for agenda 2030: Providing evidence on progress and sustainability* (pp. 51–64). International Development Evaluation Association. https://ideas-global.org/wp-content/uploads/2017/12/IDEAS-web-REV_08Dec.pdf.

Naidoo, I., & Soares, A. (2020). Lessons learned from the assessment of UNDP's institutional effectiveness jointly conducted by the independent evaluation office and the office of audit and investigation of UNDP. In M. Barrados & J. Lonsdale (Eds.), *Crossover of audit and evaluation practices: Challenges and opportunities* (pp. 182–193). Routledge.

Organisation for Economic Co-operation and Development. (2005, March 2). *Paris declaration on aid effectiveness*. https://doi.org/10.1787/9789264098084-en

Rist, R. C., & Stame, N. (2006). *From studies to streams: Managing evaluative systems*. Transaction.

Schwandt, T. (2005). The centrality of practice to evaluation. *American Journal of Evaluation*, *26*(1), 95–105. https://doi.org/10.1177/1098214004273184

Schwandt, T. (2019). Post-normal evaluation? *Evaluation*, *25*(3), 317–329. https://doi.org/10.1177/1356389019855501

The Wits School of Governance. (2020, September 8). *The future of monitoring and evaluation* [Video]. YouTube. https://www.youtube.com/watch?v=qTXhHbE3Fp4

United Nations. (2020, April). *A UN framework for the immediate socio-economic response to COVID-19*. United Nations Sustainable Development Group. https://unsdg.un.org/sites/default/files/2020-04/UN-Framework-for-the-immediate-socio-economic-response-to-COVID-19.pdf

United Nations Development Programme Independent Evaluation Office. (2020). *Leaving no one behind: Evaluation for 2030. Proceedings from the 2019 National Evaluation Capacities Conference*. United Nations Development Programme Independent

Evaluation Office. https://nec.undp.org/sites/default/files/2021-03/NEC2019_proceedings-1.pdf

United Nations Evaluation Group. (2020). *UNEG ethical guidelines for evaluations*. United Nations Evaluation Group. http://www.unevaluation.org/document/detail/2866

Universität Bern. (2018, October 17). *IPDET 2018: Brown bag 2* [Video]. YouTube. https://www.youtube.com/watch?v=71yV80jJqr4

van den Berg, R. D., Naidoo, I., & Tamondong, S. D. (Eds.). (2017). *Evaluation for agenda 2030: Providing evidence on progress and sustainability*. International Development Evaluation Association. https://ideas-global.org/wp-content/uploads//2017/12/IDEAS-web-REV_08Dec.pdf

Voccia, A. (2021). Behind the glossy cover: Six reflections on communicating evaluations. *Independent Magazine* (1). https://issuu.com/ifad_ioe/docs/independent_magazine_-_issue_1_web?fr=sZmE2MDM3MzgwMjU

Wilton Park Dialogue. (2018). *Revisiting independence, objectivity, and the critically reflective role of evaluation for the SDG era*. https://www.wiltonpark.org.uk/wp-content/uploads/2020/09/WP1600-Report.pdf

7 Do Lockdowns Work?

Evidence from the UK

Ray Pawson

Background

Across all nations, after two full years of struggle against the pandemic, there has been urgent stocktaking and considerable breast-beating about the management of the crisis. Was the response sufficient? Did it come too late? Was it allowed to relax? What mistakes were made? Was the science followed? No doubt, lessons will be learned, inquiries will be mounted, faults will be identified, and fingers will be pointed at specific decision-makers. This chapter forgoes recrimination and concentrates on explanation, trying to decipher the common challenges that undermined the management and control of the virus.

The chapter examines UK policy as the case study, although similar debates are ongoing across Europe on the persistent imprecision of virus control. There was a depressing communality. Most countries experienced an initial upsurge in infections and death, which was quelled with an initial "lockdown" – only for the virus to reassert itself. Thereafter, a succession of further lockdowns, curfews, closures, and circuit breakers were applied with initial but unsustained success. Infection rates defied wave after wave and permutation after permutation of interventions. A year on, the main prospects for virus control have shifted dramatically from controlling population behaviour to vaccination programs. This abrupt change of emphasis has left an intriguing question unanswered – in the absence of a vaccine, would it have been possible to suppress a national epidemic by social control measures alone?

Thinking about what has not happened but could have happened is the stuff of "counterfactual history" (Evans, 2014). And even though such exercises in "altered pasts and alternative futures" are sometimes dismissed as guesswork and speculation, this is the captivating and important question to be pursued here. What accounts for the relentless pattern of repeated trial and error that characterizes virus containment policy in so many countries? And, without the biological discoveries, would social suppression strategies, in the long run, have struggled to succeed?

DOI: 10.4324/9781003376316-8

I begin by acknowledging the unprecedented ferocity of biological attack, the precipitous transmissibility of the virus, and its uncanny ability to mutate genetically. But for a year, the battle against this virological enemy was conducted with what the UK Scientific Advisory Group for Emergencies (SAGE) elected to call "non-pharmaceutical interventions." It is a term that fails to do justice to the enormity of the policy response to COVID-19, which reaches down from macro-economic strategies to counteract mushrooming international debt, sweeping onward to propose comprehensive controls on every institution, organization, and service, and ends in draconian restrictions on all individual behaviour and contact. In short, the policy response under consideration here consisted of an unparalleled exercise in social control and a sociological explanation is required to account for its fragility.

The thesis here is that the UK policy response has been undermined by its own complexity. Historically, public health policy has been driven by relatively simple, linear program theories – if we introduce program X will it produce outcome Y? If we improve water supply, will it reduce diarrheal deaths? If we ban smoking in public places, will it reduce tobacco-related illness? Causal attribution is entirely different and exceedingly difficult in pandemic management, where X is a complex, adaptive, self-transformative system, aimed at a Y that encapsulates whole societies, which are themselves complex, adaptive, and self-transformative systems.

There has been a significant "turn" towards complexity and systems thinking across the social sciences in recent years (Williams, 2021) and this orientation features increasingly in policy analysis (Daviter, 2019), in implementation science (Braithwaite et al., 2018), and in the evaluation of national reforms (HM Treasury & Evaluation Task Force, 2020). The properties of complex systems have been dissected in detail (including their adaptation, emergence, unanticipated consequences, feedback loops, blockage points and structures, non-linearity, tipping points, path dependency, openness, and self-transformation). Much of the discussion of these processes has been conducted in the abstract in methodological journals and it is important to convey that that these system dynamics are in fact routine features of all social change and every social policy.

Accordingly, the chapter now begins its central task, that of articulating these wayward system dynamics as they apply to the crisis management of the UK coronavirus outbreak. Evidence is collected from a plurality of sources: the method utilized being a rapid and truncated version of "realist synthesis" (Pawson, 2006). Accordingly, the "what works" question is transformed in the expectation that the various elements of lockdown will have circumscribed impact – they will only work if implemented in particular ways, in particular communities, in particular respects, and for particular durations. Seven classic system dysfunctions are identified and for each, primary research evidence is cited on how the underlying policy assumptions become destabilized. Note that in this brief chapter it

is only possible to cover a mere handful of the countless possible examples of policy malfunctions. All illustrations refer deliberately to interventions mounted prior to the advent of the mass vaccination regime. Taken together, these perverse outcomes exemplify the remorseless challenge of complexity and so begin to explain the mixed and lurching pattern of success and failure of social containment policy across the UK (which were mirrored in many other national regimes). Significant implications follow for pandemic management and for the science of evaluation.

Modes of complexity

Interaction and emergence

The policy response, commonly referred to as "lockdown," actually consists of a very large and ever-mutating bundle of interacting programmes (i.e., hand hygiene, protective equipment, closure of shops, stadiums and schools, rules on social distancing and gatherings, restriction on travel, requirements to work from home, and many more). The impact of this medley of interventions is not simply additive but significantly interactive. Each intervention conditions the others, often in unanticipated ways. The combination of programs generates emergent effects, which may complement each other but often reduce, compete with, or displace the intended effect.

UK examples

In the first wave of the virus, the high risk of hospital acquired infection (Heneghan et al., 2020) and the urgent need for more hospital space to treat COVID patients led to a program of discharging elderly patients to care homes. This strategy succeeded in its primary, numerical aim but failed to include a testing program to accompany the transfer of patients and so displaced the problem, causing a substantial surge in care home transmission; 33 care home outbreaks in the first week of March 2020 turned into 793 by the end of the month (Public Health England, 2020a). Increasing COVID space and services in hospitals also led to substantial shortfalls in routine and planned care. Cancer services were significantly affected with a growing backlog for referrals as well as delays and cancellations of initial treatment (Macmillan Cancer Support, 2020).

A broad range of UK measures involved isolating individuals and encouraging them to remain at home. These interventions reduced virus transmission but generated a range of positive and negative emergent effects. For instance, the housebound population enabled a decrease in visits to overloaded family practitioners and to overstretched hospital Accident and Emergency Centres (McConkey & Wyatt, 2020). Working from home also decreased pollution levels, traffic accidents, and delays.

Conversely, home isolation led to measurable and damaging increases in mental health problems, domestic abuse, and educational disadvantage (Holt-Lunstad, 2020).

Even superficially simple interventions generate unintended effects. Later in the UK policy cycle, a program called "Eat Out to Help Out" subsidized restaurant-goers' bills to support restaurants struggling because of closure under wave 1 lockdown. The scheme cost £84 million and on its final day it generated a 200 per cent increase in diners over the previous year's figure (Hutton, 2020). Detractors claimed that the scheme amounted to and subsidizing of the reintroduction of the virus. There is some evidence that areas with more participating restaurants saw a notable increase in infection clusters starting around one week after the scheme launched (Fetzer, 2020).

The free rider problem

One trigger of growing resistance to lockdown stems from the activities of "free-riders." The term derives from an essay by Pareto (1935) who describes it as follows:

> If *all* individuals refrained from doing A, every individual as a member of the community would derive a certain advantage. But now if all individuals less *one* continue refraining from doing A, the community loss is very slight, whereas the one individual doing A makes a personal gain far greater than the loss that he incurs as a member of the community.

In the case of COVID-19, if one person ignores the lockdown, she or he gains from the collective effort, without having to make an individual contribution. The problem occurs when one becomes two and two becomes many. A sense of injustice amplifies if free riding becomes conspicuous and commonplace, generating a moral struggle between the "concerned" and the "unconcerned," which has a significant but once again unpredictable impact on the effectiveness of virus controls.

UK examples

The activities of free riders had a deleterious effect on UK public trust in the management of the epidemic. A longitudinal survey by Fancourt et al. (2020) charts the changes in public trust in the government's handling of the pandemic. There was a steep decrease in confidence starting in May 2020, which has never recovered. This date coincides with the discovery that Dominic Cummings, the Prime Minister's then senior advisor, had broken lockdown rules with a 500-mile round trip to a family estate. The fact that such a high-profile official had abstained from collective

responsibility ignited a torrent of media abuse – "one rule for those in charge and another rule for everyone else." Lockdown was undermined by other prominent free riders including a SAGE mathematical modeller and Scotland's chief medical officer.

After these high-profile incidents, the negative and lasting decline in public confidence was further exacerbated by crowds of anonymous free riders who gathered in parks and beaches in the early summer and at raves and house parties over the winter. But this brings us to a further disconcerting point about free riding as noted by. Let me put it carefully as it is not entirely irrational. The growing complacency of some young people incited both condemnation ("the covidiots") and empathy ("don't scapegoat the young") but, once again, it also cries out for explanation. That explanation lies in a phenomenon called "risk normalization," in which small risks become increasingly acceptable over time (Murphy, 2020). Despite year-long warnings of the savage consequences of COVID-19, most young people had no direct experience of the misery it could cause, many will have noted the limited and sporadic deterrence offered by police, and a few of them may have come across the official reports on the minute death and serious illness rates in their cohort (Bhopal et al., 2021). Putting these factors together should lead us to *expect* a significant level measure of lockdown failure.

Contextual heterogeneity

Both the transmission potential of COVID-19 and the public capacity to respond fluctuates significantly from context to context; the variations are endless. Very young children, dementia sufferers, people with physical incapacities (and the drunk and disorderly) have little capacity to obey distancing rules. Transmission varies sharply from neighbourhood to neighbourhood according to local amenities, population density, and housing stock. The socio-economically disadvantaged and many ethnic minority groups have been disproportionately affected in terms of infection rates, hospital admissions, and deaths. National lockdowns are rarely "granular" or locally sensitive enough to address every high-risk group. Even more problematic from a program theory perspective, however, is the process of "reinforcement." Policies designed to control the virus may in fact redouble the burden. They may disproportionately advantage the already advantaged groups. They may disadvantage the disadvantaged.

UK examples

National policies struggle in the face of local complexity and this is demonstrated by the limited reach of the core interventions in respect of Black, Asian and minority ethnic (BAME) communities. Prevalence, mortality, and shielding rates can be pinpointed minutely at the "ward" level and

this data shows the persistent toll of the virus on areas with high proportions of BAME residents (Otu et al., 2020). Local, "soft intelligence" identifies why these communities fared badly – for example the collapse of the local "cash-in-hand" economy, significant exposure to "fake news" media, cultural misunderstandings with providers and referral systems, stigma involved in using city-wide services, inter-generational conflict in households, social distancing problems with large families in small houses, curtailment of funeral and mourning services, and so on (West Yorkshire and Harrogate Health and Care Partnership, 2020). Broad-brush, top-down national programming can never counter such deep and locally rooted influences. There are, moreover, many other specific communities (such as retirement, student, rural, military and prison) with different but equally significant health beliefs and social mores that may hinder adherence to national policy. The same conclusion beckons – we should *expect* a degree of lockdown failure when dealing with such "micro-circuits of transmission" (Manzo, 2020).

The longstanding socio-economic health gradient (Marmot et al., 2020) is amplified by job losses due to lockdown, with disproportionate effects on those least able to protect themselves from the virus. McKinsey & Company (Allas et al., 2020) made an early estimate of unemployment risk as follows:

> The proportion of jobs at risk in elementary occupations – which employed 3.3 million people in 2019 and include jobs such as cleaners, kitchen assistants, waiters, and bar staff – is around 44 percent. In contrast, the same number for professional occupations – such as computer programmers, project managers, and accountants – is around 5 percent.

To make matters worse, job losses are then followed by perverse reinforcement effects. People not working from home or who have been furloughed must apply for benefits or seek new jobs. The available vacancies are, of course, heavily skewed under virus restrictions (Wilson, 2020). Some new jobs service the virus control measures, such as swab testers, temperature takers, and social distance facilitators. Some vacancies seek to replace workers burnt out by virus duties – opportunities in nursing and social care have never been higher. Some openings reflect changing consumer behaviour – warehouse pickers and delivery drivers. The common denominator? All these opportunities are "public facing" and thus carry elevated risks from the virus.

Implementation drift

COVID-19 containment policy throughout the world had been led, almost without exception, by central governments. Such a hierarchical approach suffers a standard problem known as implementation drift or

policy discontinuity. Centralized approaches involve long implementation chains, with the initial plans being adapted as they pass through layers of regional and local governance and then onto managers and practitioners before finding their way to the public. Such adaptation is inevitable and may have positive or negative consequences. It may lead to operational improvements, such as when intensive care units in hospitals develop "learning circles" on surge strategies to deal with the unprecedented number of critical care patients. Conversely, drift becomes a problem when implementation chains fracture and evolve along different pathways, leading to rivalry and dissent between the stakeholders in different jurisdictions. Sometimes, implementation drift is so severe it becomes implementation blockage and progress simply stalls.

UK examples

Educational policy is devolved in the four UK jurisdictions (England, Scotland, Wales, Northern Ireland) and they approached "school closure" in quite different ways. All agreed that provision of school care should remain for vulnerable children and the children of key workers, but during the first lockdown, 71 per cent of English schools remained open, compared to 34 per cent in Wales, 30 per cent in Northern Ireland, and just 24 per cent in Scotland (Sibieta & Cottell, 2020). Further drift occurred *within* each jurisdiction. Provision in England was largely organized by individual schools and even within the single locality, different rules applied on eligibility to attend, opening hours, levels of staffing, non-attendance, safeguarding responsibilities and so on, with knock-on disparities experienced in children's learning (Cattan et al., 2021). The implementation of the "same" policy meandered to different outputs and outcomes.

The biggest casualty of implementation drift was the national Test and Trace program. Notoriously, the government did not document the basis for the delivery model of this program until September 2020, long after the scheme had commenced (Department of Health and Social Care, 2020). The scheme was initiated by the civil service, bolted together rapidly, then controlled by a portmanteau of private firms including Amazon, Royal Mail, Randox, Deloitte, Sodexo, Boots, G4S, Kuenhe & Nagel, Serco, Sitel, Astra Zeneca, and GSK. A catalogue of operational shortcomings followed. In one example, the government had assumed that each case transferred to the tracing system would provide 10 to 30 contacts (the actual number was 2.4). In another example, only one in five respondents who had symptoms of COVID-19 fully self-isolated, and only one in ten respondents who had been notified they were a close contact of somebody testing positive went on to isolate for 14 days (Department of Health and Social Care, 2020). Low compliance might have been anticipated. Self-isolation is enormously challenging and "bending the rules" may seem a

rational response particularly for people who perceive themselves as low risk and have dependents and significant economic responsibilities (Office for National Statistics, 2021).

And then there is implementation blockage. The introduction of the UK's managed quarantine system stalled and spluttered for many months through an inability to coordinate the full end-to-end process from aircraft to hotel. Every aspect of the detainees' lives had to be catered to. An Institute for Government report (Mullens-Burgess & Nickson, n.d.) details how rule after rule, protocol after protocol, responsibility after responsibility was disputed between reluctant providers. International travel is one policy domain not alleviated by the vaccine roll-out: the idea of "COVID status certification" or "vaccine passports" remains hobbled by safety, security, privacy, and legal concerns (Hodgkin et al., n.d.).

Ambiguities in guidelines

COVID-19 policy imposes a mass of restrictions on normal behaviour. These restrictions are delivered in the form of guidance on which activities are permitted and which are restricted. Some ambiguity in these guidelines is inevitable, with unclear pronouncements introducing further diversity in the public response. The first uncertainty concerns the legal status of the guidance – what is law and what is merely advisory? The second opacity lies in ambiguities in the wording or phrasing of the guidance. An immense amount of effort goes into the drafting of regulations, sometimes producing scores of pages of text. But the public rarely encounter the bureaucratic texts and disparity in the everyday understanding of regulations triggers mixed levels of compliance. Guidelines cannot reach into every single aspect of human conduct. Ambiguities provide further opportunities for "bending the rules" and go on to trigger uncertain and unpredictable outcomes.

UK examples

Many government announcements blurred the distinction between law and guidance in the pandemic regulations, creating potential confusion amongst the public and the police. The key message in the original government documentation on lockdown read as follows:

> *What you can and cannot do during the national lockdown. You must stay at home. The single most important action we can take is to stay at home to protect the NHS and save lives. You should follow this guidance immediately. This is the law.*
>
> (Hickman, 2021)

Hickman goes on to point out that much of what is stated in the remainder of the document is basically public health advice. Many

exceptions to the "law" were permitted such as shopping for essentials such as food and medicine; providing support or childcare bubbles; children moving between separated parents; working where it was "unreasonable" to work from home; education, training, childcare, medical appointments, and emergencies; moving to a new house; and daily exercise. Detailed advice was offered on what was expected each of these "reasonable exceptions," but their intricacy constitutes another notable bar on compliance.

The notion of "reasonable exceptions" is a good example of a verbal ambiguity that is built into most guidelines. COVID restrictions extend to most walks of life and exemptions are always included. They are often so potentially compendious, however, that they must be captured in stock caveats such as "essential activities," "reasonable excuses," and "where necessary." Knowing exactly where to "draw the line" thus becomes problematic for officials and the public. This dilemma reached absurd proportions in what became known as the "scotch-egg-wars." During one period of restrictions in the UK, people living under Tier 2 restrictions were allowed to drink in pubs, but only if they are also consumed a "substantial meal," Ministers were badgered on whether a scotch egg constituted such a meal? Some opined yea and some stated nay, and others felt that it could, provided the order contained chips and salad.

The ambiguity of messaging becomes even more problematic when restrictions are turned on and off, and then on and off again. Analysis by The Telegraph (Dixon & Roberts, 2020) showed that there were almost 200 rule changes by the end of September 2020. In particular, UK rules on the permitted number of people allowed in bubbles and outside meetings changed quite frequently and were met with high levels of public confusion (Fancourt, 2020). Schott (2020) provides a detailed study of "graphic confusions" in the UK Government's COVID-19 official communications. One example concerns a poster explaining the rules on meetings in which the public is permitted a choice: "Your household can meet up with *one other household* indoors or outdoors" OR "You can meet up in a group of up to *six people*, outdoors only." Got that?

Novelty and routinization effects

Another class of temporal effects often noted in program evaluation concerns the changing emotional attachment to interventions over their period of operation. There is a cyclical pattern. Policies often generate an initial surge of enthusiasm with the introduction of innovative ideas (the novelty effect). There is also some pride involved in being in at the beginning of a significant initiative (the showcasing effect). These sensations often dissipate over time as programme activities fade into the background (the routinization effect). As time continues, program expectations may

become tiresome or even resented (the fatigue effect). Predicting the pace and rhythm of this self-transformation is challenging and rarely under the control of policy architects.

UK examples

Novelty and showcasing effects are clear to see. The remarkable "Clap for Carers" event, in which neighbours stood on their doorsteps banging pots and pans every Thursday at 8:00 pm, represented a significant, if "un-British," show of public affection for those battling against the virus. Some of the initial "nudge interventions" such as the double rendition of "Happy Birthday" whilst washing your hands or using funny elbow bumps also carried significant support. But, unsurprisingly, many of these first wave innovations were not sustained. Sunstein (2017), a founder of the behavioural insights approach, acknowledges that many nudges only have novelty and thus "short-term" effects.

What of the medium and long term? There is some evidence of the routinization effect, as when people seek to push back rather than withdraw under restrictions. This process is demonstrated in the significant differences in the numbers of children claiming exemptions in order to attend schools in the UK in the two periods of formal "closure." The Department for Education (2021) reported that 21 per cent of primary school pupils and 5 per cent of secondary school pupils went into school in January 2021. This compares with 4 per cent of state primary school pupils and 1 per cent from state secondaries who were in school during closures in the previous year.

Apple, Google, and the Department for Transport all collect "mobility data." A summary by the British Broadcasting Corporation (BBC) shows that trains, buses, and the Tube in London were used considerably more in later restrictions than during the first lockdown (Butcher, 2021). In March and April 2020, car use dropped to about 35 per cent of pre-pandemic levels, and, in the later Autumn lockdown, this reverted to 60 per cent of pre-pandemic levels. Workplaces also became noticeably busier in later closures. According to Google data, usage dropped by 66 per cent in lockdown 1 versus 38 per cent in lockdown two.

Polling by the Scottish Government (Director-General Education and Justice, 2021) showed that in October and November 2020, a "consistent level" of around four in ten parents of under-18-year-olds admitted to adapting COVID-19 guidance to suit their family needs. In one example 19 per cent agreed that "It's okay for my child(ren) to go into their friend's house if I don't go in with them." The main reasons provided by parents for adapting the guidance were the mental health of their children (41 per cent), followed by applying common sense (35 per cent), to help improve their own mental health (30 per cent) and to allow parents to work (26 per cent).

Whether such resistance can also be attributed to the final intervention phase of "fatigue effects" is harder to discern. And here we come to dramatic disagreement. Behavioural scientist members of SAGE were sceptical, arguing that declines in adherence had other social causes (Mitchie et al., 2020). By contrast, the World Health Organization released a long guidance document acknowledging that "pandemic fatigue is an expected and natural response to a prolonged public health crisis" (World Health Organization, 2020).

Fatigue, one suspects, lies in the mind of the perceiver. On which note, a little piece of individual testimony:

> Nearly 12 months since the country was first plunged into lockdown, this time round feels very different. We are weary, oh so weary, the kind of fatigue that hisses quietly in the background. Most of us disliked lockdown one and two, but at least with 2020's lockdowns, we had spring to look forward to and latterly Christmas – but of course the less said about that the better.
>
> (Alexander, 2021)

Exit strategies

A classic dilemma in policy evaluation concerns the sustainability of an intervention once it has ceased. Each COVID-19 intervention was time-limited on the assumption that when sufficient control of the virus was achieved, the particular restriction could be relaxed. The many and various suppression strategies were imposed and lifted at regular intervals. These transitions rarely involved the complete return to the *status quo ante*. Accordingly, more effort was often required in devising and implementing partial and prudent "exit" strategies as compared to the immediate, draconian "entry" of a full shutdown. "Unlocking" is itself a complex, adaptive, and self-transforming system. It suffers many of the same issues already identified – the interaction of components, emergent effects, rule ambiguity, drift, heterogeneous effects, and free riders. This convoluted unwinding of most restrictions adds significantly to the problem of estimating their effects.

UK examples

Closing schools, shops, theatres and so on was much simpler to implement than reopening them with capacity limitations, one-way systems, sanitizing points, and screening and booking systems. In lifting the first lockdown, government advice (Department for Business, Energy & Industrial Strategy, 2020) for retailers included: completing a COVID-19 risk assessment, cleaning more often, reminding customers and staff to wear face

coverings, ensuring social distancing, improving ventilation, taking part in the Test and Trace system, turning away people with COVID symptoms, and awareness of staff mental health needs. Additionally, some establishments were expected to keep records of all visitors, to reduce capacity, to manage queues, and to erect barriers and screens to protect staff. These are onerous expectations, especially for small concerns that make up a large proportion of the commercial sector and levels of implementation are essentially unknown.

Indirect evidence may be gleaned from a Public Health England (2020b) report which reviewed contact tracing to establish where those infected with the virus had been in the week before they were tested. "Visiting and working in supermarkets" recorded the highest weekly exposure setting of all locations, being experienced by 18.3 per cent of people testing positive. Worries that some retailers were not implementing guidance led to the introduction of fines for infringements. It is questionable whether these supplementary measures were effective. A National Police Chiefs' Council (2021) report reported that in an eight-month period, only 306 fixed penalty notice were issued to businesses in England. No data set can pinpoint exactly where people become infected, but the point here is to reemphasize that the "unlocking" of premises is an integral aspect of any "lockdown" measure, making conjectures on their effectiveness even more precarious.

At the time of writing, and thanks to the vaccination program, the final and "irreversible" lifting of restrictions is being designed. As well as the practical problems noted above on how to implement unlocking in specific locations, there are intractable macro decisions to be made on which institutions should open first. Unsurprisingly, there are fundamental disagreements on the priorities in this respect (Nabarro, 2020). The most basic idea of complexity theory is the notion of emergence (Interaction and emergence section), namely that interventions interact and may compete, limit, and displace one another. And in respect to unlocking, political and economic interests fight against health considerations as never before. As a summary statement, I cannot improve on this conclusion from the Institute for Government report (Tallow et al., 2021):

> When and how to start lifting lockdown will present the prime minister and his cabinet with some of the toughest choices they will ever have to make … At the start of the crisis, what was good for public health was also probably in the economy's long-term interests. As we move into the next phase there is a balance ministers will need to manage – they will be walking a tightrope between the risks of another surge of infections and lasting harm to the economy, people's lives, livelihoods, and prospects.

The tightrope-walking metaphor ends this discussion of the fluctuating dynamics of the UK suppression strategy. Note again that these seven

scenarios constitute just the tip of the complexity iceberg. Many other intricacies might have been cited. What are the consequences?

Conclusion

The pandemic crisis has unleashed a thousand deliberations about what could and should have been done to control the virus. As all attention and all hopes turn to biological preparations, it is important to reflect on the largest program of public surveillance and social containment outside wartime. It has left us with the never-to-be-answered question of whether these suppression strategies would have worked if left to their own devices. We can never know for certain how the UK might have fared without the advent of effective vaccinations, but this chapter contains some pointers.

In the main body of the chapter, I have attempted to illustrate the labyrinthine complexity of the mitigation strategies. The point was to show that what was done generated a muddle of contradictory forces, blocked opportunities, displaced effects, unacknowledged conditions, and unintended outcomes. Rather than seeing these as inefficiencies or conspiracies, I would argue that they were inevitable. It is what happens in complex, liberal democracies. It is what happens when single-minded objectives and simple-sounding rules are digested by a diverse population containing people who variously champion, support, comply, prevaricate, grow weary, seek exceptions, challenge, resist, and undermine those rules – and then continue to change their minds. It is what happens in countries with compressed populations, mass transportation systems, vast commercial exchange, innumerable cultural gatherings, instant and endless interaction, open public debate, and extensive worldwide interconnectedness. It is what happens when control policy is centralized with little sensitivity to local intelligence on the elusive and all-important micro-circuits of transmission (Manzo, 2020). Complex systems are perfectly designed to achieve the outcomes that emerge. Modern social life is perfectly organized in ways that multiply the microcircuits of disease transmission.

Let us now recap the counterfactual question. Let us imagine a world without AstraZeneca, Pfizer, Moderna, and the others. I have painted a picture of a continuing, protracted, and a sometimes-self-defeating struggle to combat a pandemic using lockdowns, controls, restrictions, regulations, and exhortations. After a full year of such measures in the UK, we were left with grave doubts about their sustainability. It is quite possible to identify specific blunders in the UK response – the slow and complacent initial response, the care home crisis, scandals on the supply personal protective equipment, the woeful performance of Test and Trace, and so on. These errors will no doubt take centre stage in future formal inquiries, media wrangles, and finger-pointing on the management of the pandemic. I fear that this emphasis on failures of leadership and implementation blunders may miss the point – the thesis here is that

the core problem is system malfunction. COVID-19 generated a policy response that affected every sphere of social and institutional life. The result, depicted in the main body of the chapter, was a frenzy of sticking plasters. Each measure, often perfectly valid in its own terms, interacted with other measures, producing emergent effects that were not and could not have been entirely predicted.

Does it follow that lockdown failure was inevitable? Some qualifications are in order. This is no place to begin comparative inquiry but at the time of writing the doleful picture of system failure receives some support. A BBC (2021) survey from March 2021 concludes that most European countries, especially those with stalled vaccination programs, are "once again extending lockdowns and introducing new measures." Whilst the character and numbers of lockdowns, and the rhythms of infections and deaths vary considerably from nation to nation, there are no notable instances of success. "Boot and reboot" was the European norm. And on the other side of the coin, one notes that those nations that have more nearly succeeded in virus control by lockdown have rather distinctive political, geographic, and population characteristics. Lockdown may well work in countries with authoritarian governments, compliant populations, and mass surveillance systems – though accessing uncensored evidence is difficult (Thomson & Ip, 2020). New Zealand's famed exceptionalism also has distinctive roots: geographic isolation, easy and immediate border closure, a unitary system of government, and a tiny population – the so-called team of five million (Baker et al., 2020). But for the rest of us, lockdown impacts turned out to be partial, short-lived, and indeterminate.

If the above analysis is correct, it raises a momentous question on the status and standing of evidence-based policy, and in these closing remarks I offer a distinctly brief and modest answer. In the heat of the pandemic, it was proclaimed endlessly that the UK response was "led by the science" and that "data rather than dates" would determine the choice and the timing of policy decisions. In the main body of the chapter, however, I have attempted to show how the scientifically sanctioned evidence used to guide policy was frequently undermined by further evidence gleaned in tracking complex policy outcomes. A potential paradox lurks – which evidence, which data, which provider? Who is to be believed?

Whilst the day-to-day exchanges between advisory bodies and government are, of course, invisible to all but a handful of key insiders, there is unquestionable merit in the proposition that evidence from the SAGE played a prominent role in the UK response (Clark, 2020). The public face of that response consisted of daily Downing Street briefings led by the Prime Minster and supported by the Government Chief Scientific Adviser (GCSA) and the Chief Medical Officer for England (CMO). Of the three, it was no surprise to learn which two had a grasp of the evidence. That evidence took the form of what the GCSA and CMO claimed to be the "consensus view of where we are now" (Clark, 2020) – the distillation

emerging from SAGE committee meetings with a heavy representation of physicians, virologists, immunologists, microbiologists, mathematical modellers, and epidemiologists (Thaker, 2020).

But what of the evidence presented here? A glance at the references shows that it too is a distillation – compiled in this instance from investigations carried out by a wide variety of national and local government departments, quangos, financial watchdogs, research foundations and institutes, investigative journalist, and, perhaps most significantly, by independent academics, representing disciplines such as policy evaluation, sociology, management and implementation science, and complexity and systems thinking.

It goes without saying that these "insider" and "outsider" perspectives call on different bodies of evidence but, more significantly, they carry rival understanding of the power and certainty of evidence. This relates to a schism recognized even in the pioneering days of evidence-based policy. Almost a half ago century ago, David (1975) noted scathingly that science has a taste for qualified conclusions, "on the one hand this, and on the other hand that," whilst policymakers are much more inclined to demand, "can somebody find me a one-armed scientist?" There is a remarkable echo in the UK CMO's comment to a select committee hearing:

> It is not very useful to Ministers or other decision makers to say, "There are 16 opinions. Here are all 16. Make up your mind." Part of the process is to say in a unified way, "Here is the central view."
>
> (Clark, 2020)

How has this struggle played out in UK COVID policymaking? Great caution can be seen in the many background reports utilized and submitted by SAGE: graphical projections were surrounded by confidence intervals; the possibility of measurement error was acknowledged; the erratic predictions of the mathematical modellers were protected by labelling them as "projections" or "scenarios"; recording delays were acknowledged; and tolerance was called for until "more data are collected." This steadfast restraint even resulted in injury to the English language, as in the cunning SAGE plan to uncover and counter "reasonable worse case scenarios" (surely a contradiction in adjecto). In the business of evidence compilation, methodological assiduity is the norm.

Eventually, however, prudence must be translated into policy and the broad contours of the SAGE advice have been charted earlier, namely that it consists of a vast series of recommendations on imposing and then relaxing different clusters of restrictions at different points in the evolution of the virus. I have argued that, because of complexity, there must be persistent imprecision in these recommendations. Each of the chosen measures was subject to frequent ad hoc adjustment and each package of measures unleashed unanticipated consequences, emergent properties, and displaced

effects, as described. In short, what officially sanctioned science has to offer is a large amount of guesswork – informed guesswork to be sure, but conjecture nonetheless that falls well short of certainty. Government then walks the tightrope in weaving that advice onwards into concrete policy. The inevitable result is "muddling through" and the mixture of trial and error seen in the first year of crisis management.

By this stage, however, the policy die has been cast and it needs to be justified. Once interventions have been implemented, decision-makers and the incorporated scientific elite shift subtly from caution to conviction. Alas, it must be so – policy is conjectural but can never be portrayed as such. Guidance is thus presented as scientifically sanctioned and continues to be presented by chief scientists, so carrying the imprimatur of the elite institutions. In the heat of lockdown, the prime task in daily news briefing is to bolster the decision made. When challenged, the politico-scientific establishment resort to defensive tactics, providing post hoc justifications for outcome delay and policy setback. They might suggest the measures are correct … but need better explication, so spin doctors and behavioural scientists are dispatched to redouble the advice by improving its "messaging." The measures are correct … but need more time to mature and the public is implored to keep faith, to maintain discipline, and to provide "one more push" (out of respect for health service heroes). The measures are correct … but hindered by recalcitrant people, who need to be shamed and further menaced with fines and even jail terms.

Such circumlocutions are entirely consistent with a venerable political science literature arguing that commitment to beliefs renders inflexible the attitudes of policy actors (Montpetit, 2012). In this instance, under the terrifying responsibilities of crisis management, commitment grows amongst responsible actors (such as ministers and the SAGE top table) and creates a distance with those whose beliefs differ, most especially in this instance of isolating scientists with divergent interpretations on virus control. The flow of information between disputatious parties is cut and in so doing so, science is hobbled. Real science depends unashamedly on imaginative hypotheses and guesswork. Recall Popper's (1992) assertion – "we cannot know, we can only guess." And, above all, science depends on organized scepticism (Merton, 1968). It does not depend on elite consensus and infallible evidence. Objectivity gathers in the social process, whereby independent groups of scientists compete, check, and challenge each other's interpretations (Campbell, 1988).

Without question, the scientific community has laboured ferociously in the face of the pandemic, but my charge is that politically incorporated science has feigned certitude in the face of complexity. Particular bodies of evidence have been preferred and others, including the considerable repertoire presented here, have been sidelined. The draconian restrictions carried out in the name of virus control have consequences that reach well beyond the expertise of infectious disease specialists and a plurality of

perspectives is the only inroad into complexity. Future inquires will need to look very carefully at the composition of advisory committees, ensuring that program conjectures are properly challenged before they turn into policy commitments.

It is important to ponder some "alternative futures" for the conduct of expert committees in the form of minority reports, tribunal systems, open deliberations, adversarial courts, citizens' assemblies, and so on. For discussion of these, I direct the reader to a paper by Moore and MacKenzie (2020). Practical details vary but the underlying principle is paramount:

> Creating institutions that establish norms and expectations of legitimate disagreement as part of the process of forming and communicating expert advice would make it easier for experts to stay true to their expertise and harder for politicians to hide their judgments behind the science.

References

Alexander, E. (2021, January 14). Why are people breaking the covid rules? The psychology explained. *Harper's Bazaar.* https://www.harpersbazaar.com/uk/culture/a35205563/why-are-people-breaking-the-covid-rules-psychology-explained/

Allas, T., Canal, M., & Hunt, V. (2020, May 11). *COVID-19 in the United Kingdom: Assessing jobs at risk and the impact on people and places*. McKinsey and Company. https://www.mckinsey.com/industries/public-and-social-sector/our-insights/covid-19-in-the-united-kingdom-assessing-jobs-at-risk-and-the-impact-on-people-and-places

Baker, M., Kvalsvig, A., & Verrall, A. (2020). New Zealand's COVID-19 elimination strategy. *The Medical Journal of Australia*, *213*(5), 198–200.e1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7436486/

Bhopal, S., Bagaria, J., Olabi, B., & Bhopal, R. (2021). Children and young people remain at low risk of COVID-19 mortality. *Lancet Child and Adolescent Health*, *5*(5), e12–e13. https://pubmed.ncbi.nlm.nih.gov/33713603/

Braithwaite, J., Churruca, K., Long, J. C., Ellis, L. A., & Herkes, J. (2018). When complexity science meets implementation science: A theoretical and empirical analysis of systems change. *BMC Medicine*, *16*(63). https://doi.org/10.1186/s12916-018-1057-z

British Broadcasting Corporation. (2021, June 25). *Covid: How are European countries tackling the pandemic*? https://www.bbc.co.uk/news/explainers-53640249

Butcher, B. (2021, January 13). *Lockdown: Are people breaking covid rules?* British Broadcasting Corporation. https://www.bbc.co.uk/news/55626008

Campbell, D. (1988). *Methodology and epistemology for social science*. Chicago University Press.

Cattan, S., Farquharson, C., Krutikova, S., Phimister, A., Salisbury, A., & Sevilla, A. (2021, February 19). *Inequalities in response to school closures over the course of the first COVID-19 lockdown* (Working Paper No. W21/4). Institute for Fiscal Studies. https://www.ifs.org.uk/publications/15302

Clark, G. (2020, May 18). *Letter to the Prime Minister: COVID-19 pandemic: Some lessons learned so far.* United Kingdom Parliament. https://publications.parliament.uk/pa/cm5801/cmselect/cmsctech/correspondence/200518-Chair-to-Prime-Minister-re-COVID-19-pandemic-some-lessons-learned-so-far.pdfd-early-years-settings-during-the-coronavirus-covid-19-outbreak/2021-week-3

David, E. (1975). One-armed scientists? *Science*, *189*(4204), 679. https://doi.org/10.1126/science.189.4204.679

Daviter, F. (2019). Policy analysis in the face of complexity: What kind of knowledge to tackle wicked problems? *Public Policy and Administration*, *34*(1), 62–83. https://doi.org/10.1177/0952076717733325

Department for Business, Energy & Industrial Strategy. (2020). *Working safely during coronavirus*. GOV.UK. https://www.gov.uk/guidance/working-safely-during-coronavirus-covid-19/shops-and-branches

Department for Education. (2021, January 19). *Attendance in education and early years settings during the coronavirus (COVID-19) pandemic*. GOV.UK. https://explore-education-statistics.service.gov.uk/find-statistics/attendance-in-education-and-early-years-settings-during-the-coronavirus-covid-19-outbreak

Department of Health and Social Care. (2020, December 11). *The government's approach to test and trace in England – Interim report*. National Audit Office. https://www.nao.org.uk/report/the-governments-approach-to-test-and-trace-in-england-interim-report/

Director-General Education and Justice. (2021, January 22). *Working paper: Covid-19 mitigation measures among children and young people*. Scottish Government. https://www.gov.scot/publications/working-paper-covid-19-mitigation-measures-children-young-people/pages/6/

Dixon, H., & Roberts, L. (2020, September 25). *'Does anyone know what it means?' How coronavirus rules have changed 200 times*. The Telegraph. https://www.telegraph.co.uk/news/2020/09/25/does-anyone-know-means-coronavirus-rules-have-changed-200-times/

Enria, L., Waterlow, N., Rogers, N. T., Brindle, H., Lal, S., Eggo, R. M., Lees, S., & Roberts, C. H. (2021). Trust and transparency in times of crisis: Results from an online survey during the first wave (April 2020) of the COVID-19 epidemic in the UK. *PLOS ONE*, *16*(2), e0239247. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0239247

Evans, R. (2014). *Altered pasts: Counterfactuals in history*. Little Brown.

Fancourt, D. (2020, October 23). *Half of the public don't 'fully understand' current covid-19 lockdown rules*. University College London. https://www.ucl.ac.uk/news/2020/oct/half-public-dont-fully-understand-current-covid-19-lockdown-rules

Fancourt, D., Steptoe, A., & Wright, L. (2020). The cummings effect: Politics, trust, and behaviours during the COVID-19 pandemic. *Lancet*, *396*(10249), 464–465. https://doi.org/10.1016/s0140-6736(20)31690-1

Fetzer, T. (2020). Subsidising the spread of COVID-19: Evidence from the UK's eat-out-to-help-out scheme. *The Economic Journal*, *132*(643), 1200–1217. https://doi.org/10.1093/ej/ueab074

Heneghan, C., Howdon, D., Oke, J., & Jefferson, T. (2020, October 30). *The ongoing problem of UK hospital acquired infections*. The Centre for Evidence-Based Medicine. https://www.cebm.net/covid-19/the-ongoing-problem-of-hospital-acquired-infections-across-the-uk/

Hickman, T. (2021, January 25). *The continuing misuse of guidance in response to the pandemic*. London School of Economics. https://blogs.lse.ac.uk/covid19/2021/01/25/the-continuing-misuse-of-guidance-in-response-to-the-pandemic/

HM Treasury and Evaluation Task Force. (2020, April 1). *The magenta book: HM treasury guidance on what to consider when designing an evaluation*. GOV.UK. https://www.gov.uk/government/publications/the-magenta-book

Hodgkin, R., Haddon, C., & Sasse, T. (n.d.). *Covid passports*. Institute for Government. https://www.instituteforgovernment.org.uk/explainers/vaccine-passports

Holt-Lunstad, J. (2020, June 21). *Social isolation and health*. Health Affairs. https://www.healthaffairs.org/do/10.1377/hpb20200622.253235/full/

Hutton, G. (2020, December 22). *Eat out to help out scheme* (Briefing Paper No. CBP 8978). House of Commons Library. https://researchbriefings.files.parliament.uk/documents/CBP-8978/CBP-8978.pdf

Macmillan Cancer Support. (2020, October). *The forgotten 'c'? The impact of COVID-19 on cancer care*. https://www.macmillan.org.uk/assets/forgotten-c-impact-of-covid-19-on-cancer-care.pdf

Manzo, G. (2020). Complex social networks are missing in the dominant COVID-19 epidemic models. *Sociologica*, *14*(1), 31–49. https://doi.org/10.6092/issn.1971-8853/10839

Marmot, M., Allen, J., Goldblatt, P., Herd, E., & Morrison, J. (2020). *Build back fairer: The COVID-19 Marmot review – The pandemic, socioeconomic and health inequalities in England*. Institute of Health Equity. https://www.instituteofhealthequity.org/resources-reports/build-back-fairer-the-covid-19-marmot-review/build-back-fairer-the-covid-19-marmot-review-full-report.pdf

McConkey, R., & Wyatt, S. (2020, June 30). *Exploring the fall in A&E visits during the pandemic*. The Health Foundation. https://www.health.org.uk/news-and-comment/charts-and-infographics/exploring-the-fall-in-a-e-visits-during-the-pandemic

Merton, R. (1968). *Social theory and social structure*. Free Press.

Mitchie, S., West, R., & Harvey, N. (2020, October 26). The concept of "fatigue" in tackling covid-19. *The BMJ Opinion*. https://blogs.bmj.com/bmj/2020/10/26/the-concept-of-fatigue-in-tackling-covid-19/

Montpetit, É (2012). Does holding beliefs with conviction prevent policy actors from adopting a compromising attitude? *Political Studies*, *60*(3), 621–642. https://doi.org/10.1111/j.1467-9248.2011.00930.x

Moore, A., & MacKenzie, M. (2020). Policy making during crises: How diversity and disagreement can help manage the politics of expert advice. *BMJ Clinical Research*, *371*, m4039. https://doi.org/10.1136/bmj.m4039

Mullens-Burgess, E., & Nickson, S. (n.d.). *Hotel quarantine*. Institute for Government. https://www.instituteforgovernment.org.uk/explainers/hotel-quarantine

Murphy, V. (2020). *Coronavirus is no longer seen as high risk by many – And that's undermining control measures*. The Open University Business School. http://business-school.open.ac.uk/news/coronavirus-no-longer-seen-high-risk-many

Nabarro, D. (2020). *Spotlight: Business leadership for social and economic continuity in the new COVID reality*. Skills, Systems & Synergies for Sustainable Development. https://4sd.info/covid-19-narratives/business-leadership-for-social-and-economic-continuity/

National Police Chiefs' Council. (2021, January 8). *Update on national crime trends and fixed penalty notices under covid regulations*. https://news.npcc.police.uk/releases/update-on-national-crime-trends-and-fixed-penalty-notices-issued-under-covid-regulations

Office for National Statistics. (2021, April 12). *Coronavirus and compliance with government guidance, UK: April 2021*. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronavirusandcompliancewithgovernmentguidanceuk/april2021

Otu, A., Ahinkorah, B. O., Ameyaw, E. K., Seidu, A.-A., & Yaya, S. (2020). One country, two crises: What covid-19 reveals about health inequalities among BAME communities in the United Kingdom and the sustainability of its health system? *International Journal for Equity in Health*, *19*(189). https://equityhealthj.biomedcentral.com/articles/10.1186/s12939-020-01307-z

Pareto, V. (1935). *The mind and society*. Harcourt, Brace and Company.

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

Popper, K. (1992). *The logic of scientific discovery* (5th ed.). Routledge.

Public Health England. (2020a, April 29). *COVID-19: Number of outbreaks in care homes*. GOV.UK. https://www.gov.uk/government/statistical-data-sets/covid-19-number-of-outbreaks-in-care-homes-management-information

Public Health England. (2020b, November 19). *Events and activities reported by people testing positive*. GOV.UK. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/936672/Weekly_COVID-19_and_Influenza_Surveillance_Graphs_w47.pdf

Schott, B. (2020, November 11). *Britain's covid response is graphic confusion*. Bloomberg. https://www.bloombergquint.com/gadfly/britain-s-covid-response-is-graphic-confusion

Sibieta, L., & Cottell, J. (2020, October). *Education policy responses across the UK to the pandemic*. Education Policy Institute. https://epi.org.uk/publications-and-research/education-responses-uk-pandemic/

Sunstein, C. (2017). Nudges that fail. *Behavioural Public Policy*, *1*(1), 4–25. https://doi.org/10.1017/bpp.2016.3

Tallow, G., Owen, J., Pope, T., Rutter, J., Paun, A., Nice, A., Thomas, A., Maddox, B., Haddon, C., Norris, E., White, H., & Davies, N. (2021, February 18). *Lifting the lockdown: How to approach the coronavirus exit strategy*. Institute for Government. https://www.instituteforgovernment.org.uk/sites/default/files/publications/lifting-lockdown-how-approach-coronavirus-exit-strategy_1.pdf

Thaker, P. (2020). Conflicts of interest among the UK government's covid-19 advisors. *BMJ*, *371*, m4716. https://www.bmj.com/content/371/bmj.m4716

Thomson, S., & Ip, C. (2020). COVID-19 emergency measures and the impending authoritarian pandemic. *Journal of Law and the Biosciences*, *7*(1), lsaa064. https://doi.org/10.1093/jlb/lsaa064

West Yorkshire and Harrogate Health and Care Partnership. (2020, July 15). *Addressing inequalities for BAME staff and communities*. West Yorkshire Health and Care Partnership. https://www.wyhpartnership.co.uk/application/files/4515/9713/9130/Paper_B_Developing_an_Intelligence_and_Insight_Data_Pack.pdf

Williams, M. (2021). *Realism and complexity in social science*. Routledge.

Wilson, B. (2020, August 11). *What jobs are available post-lockdown?* British Broadcasting Corporation. https://www.bbc.co.uk/news/business-53414413

World Health Organization. (2020). *Pandemic fatigue: Reinvigorating the public to prevent COVID-19*. WHO Regional Office for Europe. https://apps.who.int/iris/bitstream/handle/10665/335820/WHO-EURO-2020-1160-40906-55390-eng.pdf

8 The Role of Evaluative Information in Parliamentary Oversight of the Australian Government's Responses to the Pandemic

Peter Wilkins

Introduction

There has been extensive commentary on the accountability arrangements applicable to the actions of governments in response to the COVID-19 pandemic. Independent scrutiny options include inquiries by parliamentary committees, integrity agencies, and Royal Commissions, these having been identified through an analysis of scrutiny of the Australian Government's response to the 2008 Global Financial Crisis (GFC) (Wilkins et al., 2020).

It has been observed that in responding to the pandemic, democratic states around the world have massively expanded executive powers at the expense of oversight by legislatures and much of this through the delegation of legislative power from Parliament to the executive (Dey & Murphy, 2021).[1] The extensive scope for ministerial discretion has also been identified as a challenge to accountability (Tham, 2020).

More particularly, it has been recognised that:

> [I]n response to this complex and potentially devastating threat, parliaments around Australia have given governments unprecedented power over our day-to-day activities, travel, attendance at schools and workplaces, and welfare entitlements. They are collecting and sharing personal information, detaining individuals, and spending enormous amounts of public money.
>
> (Moulds, 2020)

Australia does not have a constitutional bill of rights and the ability of parliamentary committees to scrutinise COVID-19 laws from a human rights perspective has been questioned including"

> How do we make sure governments are keeping our rights and interests front of mind as they make these laws and give themselves these

DOI: 10.4324/9781003376316-9

> powers? How do we make sure that the "extraordinary" measures remain proportionate to the risks we face?
>
> (Moulds, 2020)[2]

Given these concerns, it is of considerable interest to understand how parliamentary committees have responded to the challenge of holding the Government to account in relation to its responses to the pandemic. Importantly, the Australian Senate (the upper house of the Australian Parliament) established the Select Committee on COVID-19 (the "Committee") on April 8, 2020, to address the Government's response to the COVID-19 pandemic and any related matters and to submit its final report no later than June 30, 2022 (Senate Select Committee on COVID-19, 2020c).

Furthermore, it is also of interest to understand the role of evaluative information in this parliamentary oversight and to consider its role alongside the views and opinions expressed by stakeholders.

Evaluative information provides the basis of evaluative judgments. It can come from many sources including evaluations, performance reporting, and performance auditing and it is important to assess its quality including reliability, validity, credibility, legitimacy, functionality, timeliness, and relevance. It has been noted that "… evaluative information that lacks these characteristics stands little chance of legitimately enhancing performance, accountability, and democratic governance" (Schwartz & Mayne, 2005, p. 1). In addition, evaluative information can arise from a rigorous analysis of obligations under law and related instruments and compliance with these requirements.

To gain insights into the role of evaluative information in parliamentary oversight, this chapter reviews the work of the Committee, which is the Parliamentary body with a broad mandate encompassing all aspects of the Government's responses to the pandemic. After an assessment of the overall work, it uses a textual analysis of the submissions to the Committee and its interim reports, along with an assessment of additional documents and hearings to identify the role played by evaluative information in the process.

The chapter includes an assessment of the specific contribution of the evaluation community to the evaluative information available to the Committee but does not examine specific evaluations or the work of individual evaluators. As the Committee is continuing its work, the chapter notes its role in bringing additional information into the public domain but does not assess whether this increased transparency enhances accountability or whether decision-making has become more evidence based as a result.

Themes arose from analysing the scrutiny of the GFC measures which are also relevant in the current context, including the recognition of

context; the need for clear programme objectives; explicit design principles; and governance, which point to but do not assess directly issues relating to the role of evaluative information (Wilkins et al., 2020).

Senate Select Committee on COVID-19

At the time of writing the Committee had received 547 submissions, 88 additional documents, and held around 50 hearings between April 23 and July 30, 2021.[3] It has to date issued two interim reports. The Committee will have presented its final report to the Parliament by June 30, 2022.

Submissions

An overview of the 547 submissions made to the Committee indicates that the vast majority conveyed the views and opinions of the individuals and organisations making the submissions but did not include evaluative information that could be used to provide an evidentiary base for findings and recommendations by the Committee. Instead, they appeared to primarily present a selective survey of attitudes about the Government's response and act as a form of unstructured consultation to provide views and ideas.

For instance, the Medical Consumers Association submitted its view that Medicare-funded psychiatric services were being provided by a "cartel," and called for an inquiry into the structure of the mental health provider sector without submitting evaluative information that supported their view.

There are relatively few submissions from professional bodies. One submission was provided by The Institute of Public Accountants, which pointed to a lack of consultation in designing welfare payments through the tax system, highlighting that the lack of consultation had caused implementation issues for practitioners and accountants. It also argued for tax reforms, comparing international wage subsidy programmes, and suggested that a guarantee-scheme be extended.

The three submissions from vaccine manufacturers funded by the Government primarily contain assertions, supported in some cases by references to other sources but did not provide full evidence supporting the statements they made. AstraZeneca provided over ten pages including brief summaries on the vaccine development, efficacy, safety, manufacturing, and regulatory approvals. Pfizer Australia and New Zealand identified some of the key issues encountered during the pandemic as well as broader issues that could improve Australia's long-term health security. It provided more than five pages of observations on the vaccine, their link to long-term health security, supply, and the national medical stockpile. CSL, an Australian biotechnology company, provided more than two pages of background detail on its manufacturing capacity and vaccine supply agreements.

One example which reinforced to the Committee the specific legal obligations of the Government was the submission by Amnesty International, set in the context of the *Universal Declaration of Human Rights* and other international human rights instruments. It recommended that human rights must be at the centre of all efforts; all decisions must be based on scientific evidence; cost should not be a barrier to accessing care; information must be available to all; and frontline workers must be protected. It also addressed issues of housing and homelessness, water and sanitation, welfare, schools, and movement restrictions. It made 17 recommendations to the Australian Government although it did not support these suggestions with an analysis of how they derived from specific laws or international treaties.

Difficulties accessing disaggregated data may partially explain the lack of quantitative analysis that underpinned submissions to the Committee. However, one exception where evaluative information and judgements were provided was the rare example of the Grattan Institute's submission, which drew some lessons for the health system derived from modelling which simulated the risks of different restriction relaxation strategies.

Additional documents

Among the 88 additional documents provided to the Committee, there were letters of correction and some additional information. These included a Treasury Ministerial Submission on Economic Impacts of Severe Acute Respiratory Syndrome (SARS) which included assessments of the exposure of the Australian economy to China and how economic activity rebounded once SARS was contained. There was also a proposal to reduce spending on JobKeeper (a programme supporting employers to retain their employees), based on modelling two examples of changes to how it was taxed.

Data was provided on payment recipient numbers, income support payments, pandemic leave disaster payments, employment, and Jobactive. This is in effect raw data and does not in itself provide any insights. It is unclear if this data was to be used by the Committee for evaluative purposes or to refer to directly in its reports.

Hearings

The Committee held around 50 hearings between April 23 and July 29, 2021.[4] Fourteen hearings were supported by one or more submissions.

The Committee's work in this regard has been summarised as:

> … calling on a range of key public officials to detail their advice to government during the pandemic. Treasury officials, for instance,

> have been quizzed about how the government's response is affecting the federal budget, health officials have been asked about the timing of travel bans and other restrictions, and programmers have been questioned about the privacy protections in the COVIDSafe app. The committee has also asked public servants about the impact of the COVID-19 response on key services such as the National Disability Insurance Scheme and on vulnerable groups such as those in remote Aboriginal communities.
>
> (Moulds, 2020)

First interim report

The Committee's first interim report is dated December 2020; it presented initial findings on a broad range of issues and made six recommendations. There were 25 interim findings (Senate Select Committee on COVID-19, 2020b, pp. xvii–xxii). It drew on 37 public hearings, 505 written submissions, and answers to questions on notice (Senate Select Committee on COVID-19, 2020b, p. xi). The 37 hearing days listed (Senate Select Committee on COVID-19, 2020b, pp. 189–220) took place between April 23 and November 26, 2020, and included a large number and variety of parties and witnesses.

The main body of the report is 122 pages long and included chapters on:

- Preparation and initial response;
- health response;
- economic response;
- national governance; and
- looking ahead.

It contained 57 footnotes detailing the sources for statements made which included the Government, submissions, press conferences, and media coverage.

There was also a dissenting report by Coalition Senators totalling 45 pages which referred to the preceding material as a majority report and additional comments from Australian Greens Senators of a further 13 pages. While this indicates that there are differences along political party lines, it has been argued that minority reports can be categorised as primarily policy, political, malpractice, malfeasance, or evidential in nature (Aliferis & Mackay, 2021).

The interim report made little use of words linked to evaluation, but interesting use of the term "evidence"; it appears to be used as meaning any information that was submitted or said during the inquiry, without any formal assessment of its quality. For instance, it states that the report "… is principally based on evidence provided to the Committee via 37 public hearings, 505 written submissions and answers to questions

on notice provided by government departments, agencies, and other witnesses" (Senate Select Committee on COVID-19, 2020b, p. xi).

However, it is evident in the report's content that evidence was assembled and weighed, providing implicit acceptance of the quality of what is presented. For instance, in commenting on the Australian Government's early response to COVID-19, it observed that "[t]hroughout March [2020], the Prime Minister gave mixed messages on the role and importance of social distancing in reducing community transmission and, at times, appeared to cast doubt on the need for such measures," going on to quote the Prime Minister as having said "… that people could 'go about their daily business' and that he was 'looking forward to going to places of mass gatherings such as the football,'" and then adding that "Dr Norman Swan gave evidence to the Committee that this was 'the dominant message from governments, particularly the Federal Government' prior to 16 March" (Senate Select Committee on COVID-19, 2020b, p. 35).

The word "research" appeared in various contexts, mostly referring to statements and claims made during hearings. For instance:

> On 25 June Dr MacIntyre, who has conducted "the largest body of clinical research on face masks and respirators in the world," told the committee that the use of face masks as a tool to reduce the transmission of COVID-19 was "cheap, effective and low risk," as well as "a no brainer."
>
> (Senate Select Committee on COVID-19, 2020b, p. 42)

There were a few instances where research is cited directly, for instance:

> [A] report by the Australian Housing and Urban Research Institute in November noted that for state and territory governments, "social housing has featured as a key plank of the economic recovery platform," but that "there has been no new direct allocation of funding for social housing by the Australian Government, which contrasts with the Global Financial Crisis, where $5.2 billion ($6.5 billion in 2020 dollars) was allocated to the Social Housing Initiative."
>
> (Senate Select Committee on COVID-19, 2020b, p. 104)

A specific assessment of the information used in the report to support the six recommendations in the report[5] (see Table 8.1) reinforces a view that evaluative information did not play a central role in their formulation.

Two of the recommendations concern transparency (1 and 6), two were about review-type activity (2 and 4), one requested an increase a level of support (5) and one proposed the establishment of a new body (3). In the following sections, we assess the evaluative information for each of these groups.

Table 8.1 Supporting information for the six recommendations of the Committee's first interim report.

Recommendation and Page No.	*Subject Matter*	*Summary*	*Supporting Information*	*Comments*
1. Page 39	Lack of transparency behind key medical decisions.	Publish all previous and future Committee minutes.	The advice provided to National Cabinet by the AHPPC is unnecessarily secretive. No minutes of meetings have been made publicly available since 26 February. Several hundred meetings were held in the first eight months of the pandemic and only 65 statements have been released to the public.	The report highlights a lack of clarity about the national strategy and advice by the AHPPC would help to clarify the basis of statements by politicians. Disputed by Government Senators.
2. Page 43	COVIDSafe had under-delivered and had only been of limited effectiveness.	Commission an independent review into the COVIDSafe app.	The Government had announced that use of the app was a condition to gradually re-open the country.	Evidence of technical problems and limited successful use of the app for contact tracing.
3. Page 49	Australia was the only Organisation for Economic Co-operation and Development (OECD) member without a Centre for Disease Control (CDC).	Establish an Australian Centre for Disease Control.	Many submissions argued for establishment of some type of CDC and made the case that it could have addressed some of the concerns identified by the Committee.	The Committee noted that the model needs were determined in close consultation with key stakeholders. Government Senators did not comment on this recommendation.

(*Continued*)

Table 8.1 Supporting information for the six recommendations of the Committee's first interim report. (*Continued*)

Recommendation and Page No.	*Subject Matter*	*Summary*	*Supporting Information*	*Comments*
4. Page 74	A Supplement which effectively doubled the JobSeeker payment had been introduced.	To monitor the economic impact of reducing the Coronavirus Supplement and report back to the Senate.	The supplement was extended twice at reduced rates and was due to finish on March 31, 2021.	When asked the Minister was unable to answer the Committee's questions on the economic impact of reducing the Supplement. Government Senators rejected the assertion that the JobSeeker payment was inadequate.
5. Page 82	Linked to Recommendation 4.	Permanently raise the rate of JobSeeker payment.	The Committee cited wide acknowledgement that the permanent rate of JobSeeker payment was inadequate, often forcing recipients to live well below the poverty line.	The Committee cited university research and argument for a sustained increase to the JobSeeker rate. Government Senators rejected that the JobSeeker payment was inadequate.
6. Page 117	Lack of transparency.	Make public all reports of the National COVID-19 Commission Advisory Board (NCCAB)[6] and any declarations conflicts of interest.	Interim finding: had access to cabinet documents without commensurate accountability, had not released any work publicly and failed to demonstrate how conflicts of interest are managed for commissioners.	The report states that the origins of the Board are unclear and that there was significant concern about Commissioners' actual, perceived, or potential conflicts of interest. This was disputed by Government Senators.

Transparency

The first recommendation on transparency (recommendation 1) related to the publication of the minutes of the Australian Health Protection Principal Committee (AHPPC) "… to provide the public with access to the medical advice behind all decisions affecting the community's safety, livelihoods and personal freedoms" (Senate Select Committee on COVID-19, 2020b, p. 39).[7]

In support of this recommendation, the Committee noted that it was "unclear exactly when the government first became aware of an elevated risk associated with travellers returning from the USA and Europe" (Senate Select Committee on COVID-19, 2020b, p. 16). It was equally unclear when the government became aware of risks from some countries which had limited ability to ascertain transmission levels. This lack of clarity can be attributed to the government's refusal to provide the Committee with access to key documents such as AHPPC minutes which would confirm that information.

As a specific example, in May 2020, the Department of Health (DoH) clarified that the National Cabinet had endorsed a strategy of suppression with the potential for elimination. Further information was provided on July 24 when the AHPPC announced that "the goal for Australia is to have no community transmission of COVID-19, strengthening the current suppression strategy" (Senate Select Committee on COVID-19, 2020b, p. 34).

However, the DoH refused to provide any information about whether the decision was made based on AHPPC advice or which materials were relied on in making that decision. In light of this refusal to provide information, the Committee concluded there was no evidence that the strategy articulated by the Prime Minister on September 4 was the same as the strategy adopted early on by the government (Senate Select Committee on COVID-19, 2020b, p. 35). In making the case that the public should have access to the medical advice behind Government decisions, the report observes that "[e]xpert witnesses presented compelling evidence to the Committee that AHPPC advice should be publicly available" (Senate Select Committee on COVID-19, 2020b, p. 41) and quoting the statement of a witness supporting this view.

However, Government Senators noted that:

> The government has continually taken a strong and decisive approach in responding to COVID-19. This has been informed by the latest technical and scientific advice from the Australian Health Protection Principal Committee (AHPPC) and its Standing Committees, in particular the Public Health Laboratory Network and the Communicable Diseases Network Australia.
>
> (Senate Select Committee on COVID-19, 2020b, p. 125)

The second transparency recommendation (recommendation 6) asked that all reports of the National COVID-19 Commission Advisory Board (NCCAB) be made public, together with all declarations of actual and perceived conflicts of interest made by commissioners (Senate Select Committee on COVID-19, 2020b, p. 117).

The NCCAB was endorsed by all first ministers in March 2020 as a non-statutory body with its commissioners appointed by the Prime Minister. The Committee made an interim finding that the NCCAB lacked sufficient transparency and had access to Cabinet documents without commensurate accountability, had not released any work publicly, and failed to demonstrate how conflicts of interest are managed for commissioners (Senate Select Committee on COVID-19, 2020b, p. 117).

Government Senators disputed the interim finding, arguing that the NCCAB "… sits within the Department of the Prime Minister and Cabinet and is bound by the usual governance protocols and processes, including in relation to procurement," that it had appeared before the Committee on three occasions, and that it "… is subject to other normal transparency mechanisms of other government agencies, including freedom of information requests and public reporting of contracts" (Senate Select Committee on COVID-19, 2020b, p. 166).

Review-type activity

The first recommendation calling for a review (recommendation 2) demanded an independent review into expenditure on and design of the COVIDSafe app (Senate Select Committee on COVID-19, 2020b, p. 43).

On May 1, 2020, the Government had announced that the use of the COVIDSafe app, which had launched on April 26, was a condition to gradually re-open the country. The app was designed to assist contact tracing by being voluntarily installed on mobile phones, and to record and store data when a phone had been within close proximity of another for a specified period, requiring both individuals to have the app installed and open.[8] The interim finding concluded the app had significantly underdelivered and that its effectiveness was limited.

Government Senators recommended that "[o]ngoing improvements are made to the app through regular updates" and that "[t]he Government continues to support and encourage state and territory governments to use the app to supplement their contact tracing process" (Senate Select Committee on COVID-19, 2020b, p. 148).

The second recommendation calling for a review-type activity (recommendation 4) urged the Government to monitor the economic impact of reducing the Coronavirus Supplement and report back to the Senate with any data on its impact. The Committee cited university research and accompanying arguments for a sustained increase to the JobSeeker rate.

A Coronavirus Supplement, which effectively doubled the JobSeeker payment, was due to end on September 24, 2020. But on July 21, the Government announced that it had decided to extend the payment at a reduced rate until December 31, 2020. When asked in August, the minister responsible was unable to answer the Committee's questions on the economic impact of reducing the Supplement. In November, the Government announced that the Supplement would be extended at a further reduced rate until March 31, 2020.

Increased level of support

Linked to recommendation 4, Committee recommended (recommendation 5) that the Government permanently raise the rate of the JobSeeker payment at the Mid-Year Economic and Fiscal Outlook or in the 2021–2022 Budget.

The report cited wide acknowledgement "… for years, by business and community stakeholders alike that the permanent rate of JobSeeker at $40 a day is totally inadequate, often forcing recipients to live well below the poverty line." It referred to university research showing "… that the introduction of the Coronavirus Supplement and the JobKeeper payment 'reduced measures of poverty and housing stress, with both now below what they were prior to COVID-19'" (Senate Select Committee on COVID-19, 2020b, p. 85). This research is another rare example of evaluative information considered and cited by the Committee. The analysis combined two sources of data: (1) an impact monitoring survey, the only publicly available longitudinal survey in Australia with information tracking individuals before the spread of COVID-19 and through the receipt of JobSeeker and JobKeeper payments; and (2) a microsimulation model of the tax and transfer system (Phillips et al., 2020, p. iii).

It also cited argument from the Grattan Institute "… for a sustained increase to the JobSeeker rate to around $750 a fortnight on the basis that the 'Coronavirus Supplement has been very important for sustaining spending and incomes through this period' " (Senate Select Committee on COVID-19, 2020b, p. 85).

Government Senators rejected the claims that the JobSeeker payment was inadequate, and that the Government was withdrawing fiscal support too early (Senate Select Committee on COVID-19, 2020b, p. 156). They noted that "[i]t is the responsibility of the government to ensure our social security and welfare system is sustainable into the future, so that it can continue to provide support to those most in need" and that "[t]he emergency income support measures the Commonwealth government put in place at the outset of the coronavirus pandemic were always targeted, temporary and scalable" (Senate Select Committee on COVID-19, 2020b, p. 157).

Establish a new body

The Committee recommended (recommendation 3) that the Government should establish an Australian Centre for Disease Control (CDC) to improve Australia's pandemic preparedness, operational response capacity, and communication across different levels of government. It noted that Australia was the only OECD member without a CDC. Many submissions argued for the establishment of some type of CDC; the Committee accepted these arguments and observed that the best model needed to be determined in close consultation with key stakeholders including those in the aged care sector, the states, and the territories (Senate Select Committee on COVID-19, 2020b, p. 51).

Second interim report

The Committee's second interim report, dated February 2021, deals with claims of public interest immunity, where information disclosure can be avoided on the grounds it may be prejudicial to the public interest, and which were received by the Committee from several ministers during its examination of the Australian Government's response to the pandemic.

The Committee observed that:

> These claims have compromised the committee's ability to scrutinise government decisions with a profound impact on lives of Australians. The committee is concerned that they reflect a pattern of conduct in which the government has wilfully obstructed access to information that is crucial for the committee's inquiry.
>
> (Senate Select Committee on COVID-19, 2021, p. 1)

The Committee made seven recommendations to the Senate requesting it pass resolutions requiring access to the documents sought by the Committee, explaining its reasoning in each case why it did not accept the minister's claims of public interest immunity. The documents are in many cases answers to written questions asked by Committee members.

A section of the report provides Government Senators' additional comments, which highlights the extensive information provided to the Committee, including nearly 2000 answers to questions on notice, describing it as "… a remarkable feat of cooperation and transparency especially when considering they did so while managing the day to day fight against a once in a generation global pandemic and associated economic crisis," then going on to state that "[t]he relatively few disagreements between the Committee and the government about a handful of public interest immunity claims should be viewed in this light" (Senate Select Committee on COVID-19, 2021, p. 19).

In the context of this chapter, the difficulty the Committee encountered in obtaining information it saw as necessary and the fact that it pressed further to obtain it to progress its work all point to the high value it placed on information as evidence.

Discussion

Access to information by the Committee and more generally by the public has been a central issue. As early as June 2020, the Committee wrote about:

> Concerns with the forthrightness of departments and agencies in answers to oral and written questions. A tendency to refrain from providing full and complete responses to the Committee is particularly evident in answers about information that may have a connection to cabinet.
>
> (Senate Select Committee on COVID-19, 2020a)

and it reminded witnesses "… that there is no category of documents which the Senate has accepted is immune from production" (Senate Select Committee on COVID-19, 2020a). However, no particular mention was made about the importance of providing evaluative information.

Given the paucity of evaluative information submitted to the Committee, and the apparent indifference to requiring this information, the chapter turns next to consider two sources of evaluative information that could have supplemented those provided through the Committee's normal approaches to obtaining input as detailed above: performance audits and the evaluation community.

The first source of evaluative information is the work of the Australian National Audit Office (ANAO); it is notable that while there are relevant performance audits, the role of this evaluative information is not evident in the Committee's processes. The second source is the evaluation community; it is also remarkable that the work of evaluators in Australia is similarly absent. The material that is available to the Committee and possible reasons for its absence in the Committee's processes are presented below.

Audit to the rescue?

A notable absence from the submissions, hearings, and interim reports is any input by the ANAO. The ANAO website identifies that on March 31, 2020, a Senator for the Australian Capital Territory (ACT) and shadow minister (who was soon thereafter to take on the role of chair of the Committee on COVID-19) wrote to the Auditor-General

requesting that the ANAO develop an audit programme related to the Australian Government's economic response to COVID-19, the reasons including:

> … the unprecedented amount of taxpayer funds proposed to be expended over the next two financial years, along with the novel arrangements and extra powers being provided to Ministers over coming months, Labor believes there are good reasons for you to audit the implementation and ongoing performance of the Australian Government's economic response to COVID-19.
>
> (Gallagher, 2020)

The Auditor-General replied on April 23, 2020, that he intended "… to develop and publish an audit program of the government's COVID-19 response … The audit programme will focus on providing Parliament with transparency and assurance on management of the response" and recognised at the time of his reply that the Senate had established the Committee on COVID-19. He provided a specific reference to the "… quality of evidence developed within the public sector to inform the Minister's decisions" in a particular area (Hehir, 2020).

An ANAO newsletter article released on April 16, 2020 examined the rapid implementation of Australian Government initiatives, emphasising the importance of effective implementation to achieving government policy goals, and identifying key lessons which are likely to have wider applicability to the public service's support of the national COVID-19 pandemic response, the most relevant learning here being to "establish fit for purpose governance and planning arrangements" (Australian National Audit Office [ANAO], 2020).

The ANAO has also provided a COVID-19 multi-year audit strategy. The strategy encompassed performance and financial audits, and assurance reviews of payment advances to the finance minister. The performance audits were to be delivered in three phases.

The first phase examines "how the audited entities: manage and respond to risks related to rapid development and implementation of COVID-19 measures and the impact on business as usual activities and controls; and communicate and implement revised risk tolerances across the business"; five audits have been tabled including two related to aspects of the National Medical Stockpile (ANAO, n.d.).

The second phase focuses on "the three main stages of program delivery: policy design; implementation; and performance assessment, evaluation and dissemination of lessons learnt" together with an ongoing focus on risk management and as of July 30, 2021 four audits are listed as in progress, including three related to travel, and potentially a further five, including one on the COVIDSafe app (ANAO, n.d.).

The third phase will review the outcomes of the Government's COVID-19 response.

While not specifically recognising the contribution the ANAO could make to the work of the Committee, the strategy recognises that it "is designed to: respond to the interests and priorities of the Parliament of Australia; and provide a balanced program of activity that is informed by risk, and that promotes accountability, transparency and improvements to public administration" (ANAO, 2020).

The Auditor-General has a formal relationship with the Joint Committee of Public Accounts and Audit (JCPAA), which has the role of consulting with parliamentary committees to determine the audit priorities of the Parliament. It communicated these to the Auditor-General on May 29, 2020, following consultations regarding COVID-19 (Joint Committee of Public Accounts and Audit [JCPAA], 2020, p. 2). The Auditor-General's reports are considered by the JCPAA, and this may at least partially explain why the reports have not, at least to date, been recognised formally by the Committee. Nevertheless, it is evident that the ANAO could provide substantial input of evaluative information to underpin the Committee's work.

State and Territory Auditors General can provide evaluative information about impacts and management in their individual jurisdictions. For instance, the Auditor General for Western Australia has outlined a strategic approach to auditing the State's COVID-19 response over a time horizon of the next few years and included as its possible purposes "providing Parliament with assurance and transparency over major activities and spending linked to the COVID-19 response and whether this is delivering intended outcomes" and "evaluating the quality and timeliness of entity advice to Government" (Office of the Auditor General Western Australia, n.d.-a). Three COVID-19 specific reports to date relate to governance and assessment arrangements for a relief fund (Morrissey et al., 2020b); preparedness for the provision of pathology testing during the pandemic (this being a limited assurance review, which is not an audit) (Morrissey et al., 2020a); and the integrity of the contact registration app (Morrissey et al., 2021). The list of performance audits that have already commenced includes one COVID-19 specific report on the roll-out of stimulus initiatives (Office of the Auditor General Western Australia, n.d.-b).

However, it is not evident that there has been any coordination between State and Territory Auditors General to permit a consolidation of their findings enabling meaningful comparisons. This has been the case in rare situations, including when seven Auditors Generals concurrently investigated the implementation of a national agreement on homelessness, and a report by the Commonwealth Auditor General which included an overview that drew together common themes from five of the other reports that were already released by four State Auditor Generals and the Auditor General of the Northern Territory (ANAO, 2013, p. 69).

Evaluation to the rescue?

It is notable that the Australian Evaluation Society (AES) did not make a submission and neither did public administration and public policy bodies that might have argued for a greater role for evaluative information in the inquiry.

An AES committee issued a "Statement on evaluation during the pandemic" in June 2020 which could have been readily converted to a committee submission. The statement highlights the role that evidence and evaluation can play, and comments that the "evidence-informed approach that has served us well to-date remains equally critical going forward" and expresses the belief "that sound data collection and analysis should be built into the establishment of any new or adapted initiatives to maximise the value of evaluation" (Australian Evaluation Society [AES] Relationships Committee, 2020).

The AES has posted on its website a range of resources, including "tips, tricks, and resources for evaluators, especially related to changing work practices" (AES, n.d.-a) and they cover issues such as the challenges posed by travel restrictions to field-based assessments; developing methods to understand the complexity of the COVID-19 pandemic; and learning from evaluations of responses to past global crises such as the 2008 GFC and the 2006 avian influenza outbreak.

An early understanding of the significance of the pandemic for evaluation was demonstrated in an Editorial Foreword in the June 2020 issue of the Evaluation Journal of Australasia (EJA) with comments that included:

> While the focus of this special issue is on "values" in evaluation, it seems remiss not to acknowledge the swift and global spread of the COVID-19 virus since December last year. In light of the health, economic, ethical, and social challenges arising from the pandemic, discussions about "values" seem ever more critical in the context of these shared but localised challenges. … While evaluation may not be at the top of the "urgent expenditure" list, as governments and other organisations begin implementing recovery efforts, attention is also turning to how we make sense of various pandemic responses (globally, nationally, locally). In months and – likely – years to come, there will be ample opportunity to analyse different political, economic, and social decisions, and the impacts of those decisions. Economists, auditors, investigators, ethicists, and others may be first to review emergency management and other efforts. What opportunities will be afforded to evaluators?
>
> (Gould, 2020, p. 62)

The AES has supported an open call for COVID-19 related papers by its journal, the EJA:

> We are calling for papers on evaluation and COVID-19, to better understand responses to the pandemic – locally, nationally, or globally – and

> to illuminate the unique insights that evaluation brings. Papers may examine pandemic responses by government, private organisations, the not for profit sector, or the community, or consider conflicting views as well as the ethical challenges that underpin these conflicts … [and] consider how the pandemic has influenced evaluative work and the evaluation sector, both in practical or immediate terms, and more broadly, by way of the authorising environments for evaluation.
>
> (AES, n.d.-b)

It adds that topics of value could include evaluation insights from:

- Health, disability, aged care, and other social care sectors;
- education, tertiary, and skills sectors;
- impacts on international development monitoring and evaluation;
- evaluation of emergency management, government, or community organisational responses to the pandemic, and pandemic recovery;
- ethical and equity issues;
- economic and funding responses, including socio-economic impacts;
- digital technologies, disruption, and innovation in evaluation practice;
- methodological or theoretical impacts; and
- impacts of COVID-19 on the evaluation sector, including future impacts and implications for practice (AES, n.d.-b).

To date, the EJA has published one article that specifically focuses on an aspect of COVID-19, proposing an approach to assessing the impact of Australia's emergency response to the pandemic across its states and territories (Buck, 2020). An editorial in the journal notes that it "will continue to receive submissions for some time yet" and that "over the next year, the EJA quarterly issues will have an array of interesting papers highlighting evaluation during the COVID-19 era" (Rossingh, 2021, p. 68).

The timing of the publication of these articles indicates there are likely to be considerable delays before peer reviewed publications which present evaluative information can become available for use by parliamentary committees. More generally than the EJA timetable, at a time when there are calls for "real-time evaluation" it seems that peer reviewed journals are not well equipped to provide timely input into the deliberations of parliamentary committees in the time of a pandemic or similar emergencies.

While the collective work of the evaluation community does not indicate that evaluators in Australia have yet been called on or been able to create opportunities for the provision of evaluative information to parliamentary committees, it is evident that evaluation could provide important findings as a foundation for structured learning about the comparative impact of state and territory level pandemic response measures. For example, a methodical approach to identify such comparative impacts has been

proposed in relation to "… virus transmission and related COVID-19 morbidity and mortality" (Buck, 2020, p. 199). It is noted that "… it is critical to ensure that variables accounting for differences in implementation processes are explicitly and adequately reflected in the theoretical models on which this impact analysis is based" (Buck, 2020, p. 208).

Important contributions to evaluation can come from outside the recognised evaluation community as well. For instance, an assessment of the governance arrangements in response to COVID-19 in Australia included commentary on evaluating crisis responses and challenges including the role of political judgments which create uneven social impacts, and the difficulty of establishing benchmarks. The approach adopted for a desktop evaluation of the acute phase of Australia's response is based on the crisis context, the improvisation of plans and responses, response leadership, and knowing when to act and by how much, as these "… all influence the outcome of crisis management" (Bromfield, 2021, p. 6). The author concludes that:

> What the governments of Australia might learn from the vaccine rollout is that authoritative policy tools like lockdowns, restrictions and enforcements may work as temporary means to limit the spread of disease. But delivering a positive pandemic measure, like a vaccine rollout, is unlikely to be achievable via authoritative regulation and coercion in a liberal-democratic context and instead requires greater engagement with crisis communication and whole of government organisational delivery. This finding might be useful in future debates about the design of Australian crisis responses.

and cautions that:

> [T]his picture of conflicting evidence will complicate the calls to hold leaders and institutions to account and will likely blunt attempts to reform policies and processes to improve crisis responses in future. Students of crisis responses and evaluations should be aware of these dynamics when performing their own evaluations of the Australian COVID-19 response or crises more generally.
>
> (Bromfield, 2021, p. 12)

The particular challenges that pandemic-related evaluations face need to be recognised and may go some way to explaining why the evaluation community has not been more visible in this context to date. However, the difficulties encountered by the Committee in accessing information suggest that where there is evaluative information outside the control of departments and agencies, it may be particularly appreciated by the Committee.

Conclusion

It has been observed that committees of the Parliament "… can attract media attention, call ministers and bureaucrats to account and generally mobilise the resources of the state in ways available to few other actors … the Senate is ideally positioned as a potential committee house" (Marsh, 2006).

As a comment on the work of federal COVID-19 committees, the question is posed "whether they were up to the job of providing a meaningful check on executive power and scrutinising the impact of government actions on rights"; the short answer is that "the proof will be in the pudding." It was commented that the "… clearest way to measure the impact of these committees is to look for changes in the laws and policies themselves, such as changes that better balance individual freedoms against collective healthcare or economic interests" (Moulds, 2020).

The inquiry by the Committee on COVID-19 is continuing and it is too early to reach a final view on how it accesses and uses evaluative information. It is also premature to assess whether it is making a worthwhile contribution, but one promising sign is that:

> … the government responded to concerns about a COVIDSafe app by introducing draft legislation to limit how information is collected, stored, and used. The committee process has also helped to uncover who is missing out on social welfare packages and to highlight the complex impact social distancing measures are having on different business and service sectors.
>
> (Moulds, 2020)

The Commonwealth Auditor-General has already provided evaluative information to the Parliament and while it is formally reviewed by a different committee, it would be expected that the Senate Committee would make use of the information as it becomes available. There is also potential for coordination between State and Territory Auditors General to enable a consolidation of their findings by the Commonwealth Auditor-General to assist the Senate Committee; the Senate Committee could also coordinate its work with that of equivalent State and Territory committees.

It is not yet evident whether evaluators in Australia have been called on to contribute evaluative information that can inform the many decisions and judgments being made in response to the COVID-19 pandemic. While not having the resources of professional bodies such as those representing accountants, the AES may not have had a previous focus on contributing to Parliamentary processes. It could have fairly readily made the case for supporting an important role for evaluation to help guide the Committee's thinking regarding the role of evidence alongside views and opinions. However, internationally, it appears that scheduled evaluations in the development sphere may need to be

postponed or cancelled (Independent Evaluation Office/United Nations Development Programme & Organisation for Economic Co-operation and Development/Development Assistance Committee, 2020). This reinforces the need for the agility and insight that evaluation must "repurpose and adapt its focus and approach" to ensure its utility (United Nations Office of Internal Oversight Services, 2020, p. 8).

Compared with the long-standing and legislation-based relationship between the Parliament and the Auditor-General, in the absence of statutory independence, the evaluation community would need to make additional efforts to have its role and potential contribution respected, valued, and used.

Notes

1 A Senate Committee has inquired into the exemption of delegated legislation from Parliamentary oversight. It included legislation in times of emergency such as responses to the pandemic. In June 2021, the Senate adopted three Committee recommendations seeking to reassert control over executive lawmaking. See, Senate Select Committee on COVID-19 (n.d.-a).
2 The Parliamentary Joint Committee on Human Rights has issued a number of reports which scrutinise pandemic-related legislation for human rights compatibility. See, Parliamentary Joint Committee on Human Rights (n.d.).
3 The Committee called for submissions from the public and in September 2020 indicated that it would consider submissions provided throughout the inquiry. The submissions accepted are available on the Committee website (see, Senate Select Committee on COVID-19, n.d.-c).
4 Transcripts of the hearings are available on the Committee's website (see, Senate Select Committee on COVID-19, n.d.-b).
5 The interim findings are presented as part of an Executive Summary but are not mentioned specifically in the body of the report. It was therefore not practical to use these to assess the nature of the supporting evidence.
6 National COVID-19 Commission Advisory Board (NCCAB), originally named the National COVID-19 Coordination Commission (NCCC).
7 The Australian Government provides general information related to the pandemic on its website (see, Australian Government, n.d.) and from a health perspective (see, Department of Health, n.d.-a). While health-related news and media items related to the pandemic can be reviewed online (see, Department of Health, n.d.-b), the Government does not provide an accessible platform to view in real time the applicable statutory instruments. This is provided by at least one private law firm (see, Clayton UTZ, 2020). The Government does provide a Register of Legislation declarations made under individual statutes (see, Office of Parliamentary Counsel, n.d.).
8 See, Department of Health (2021).

References

Aliferis, J., & Mackay, A. (2021). Disrupting consensus in parliamentary committees? Minority reports and a taxonomy for classifying them. *Australasian Parliamentary Review*, *36*(1), 61–86. https://www.aspg.org.au/wp-content/uploads/2021/07/Disrupting-Consensus.pdf

Australian Evaluation Society Relationships Committee. (2020, June 29). Statement on evaluation during the pandemic. *Australian Evaluation Society*. https://www.aes.asn.au/aes-blog/statement-on-evaluation-during-the-pandemic

Australian Evaluation Society. (n.d.-a). *COVID-19 resources – Tips and tricks*. https://www.aes.asn.au/covid-19-resources

Australian Evaluation Society. (n.d.-b). *EJA open call for papers – COVID-19*. https://www.aes.asn.au/eja/eja-covid19

Australian Government. (n.d.). *COVID-19 vaccines*. https://www.australia.gov.au/covid19vaccines

Australian National Audit Office. (n.d.). *ANAO COVID-19 multi-year audit strategy*. https://www.anao.gov.au/work-program/covid-19

Australian National Audit Office. (2013). *Implementation of the national partnership agreement on homelessness*. https://www.anao.gov.au/sites/default/files/201213%20Audit%20Report%20No%2031.pdf

Australian National Audit Office. (2020, April 16). *Audit insights: Rapid implementation of Australian government initiatives*. https://www.anao.gov.au/work/audit-insights/rapid-implementation-australian-government-initiatives

Bromfield, N. (2021, October). *Evaluating governance of the COVID-19 crisis in Australia* (Preprint). ResearchGate. https://www.researchgate.net/publication/355367696_Governance_of_the_COVID19_crisis_in_Australia_public_policy_during_crisis

Buck, T. (2020). Analysing the impact of the Australian health sector emergency response plan for novel coronavirus (COVID-19): A proposed approach. *Evaluation Journal of Australasia*, *20*(4), 197–211. https://doi.org/10.1177%2F1035719X20969986

Clayton UTZ. (2020, August 7). *Government powers and directives*. Clayton UTZ. https://www.claytonutz.com/covid-19-response/government-powers-and-directives

Department of Health. (n.d.-a). *News and media*. Australian Government. https://www.health.gov.au/news?f%5B0%5D=field_related_health_topics%3A18895

Department of Health. (n.d.-b). *Our response to the pandemic*. Australian Government. https://www.health.gov.au/health-alerts/covid-19/government-response

Department of Health. (2021, November 8). *COVIDSafe app*. Australian Government. https://www.health.gov.au/resources/apps-and-tools/covidsafe-app#how-covidsafe-works

Dey, P., & Murphy, J. (2021, February 1). *Accountable lawmaking: Delegated legislation & parliamentary oversight during the pandemic* (Policy Brief No. 9). University of Melbourne. https://apo.org.au/sites/default/files/resource-files/2021-02/apo-nid310793.pdf

Gallagher, K. (2020, March 31). *Letter to Grant Hehir, Auditor-General for Australia*. Australian National Audit Office. https://www.anao.gov.au/sites/default/files/request_for_audit_senator_katy_gallagher_31_march_2020.pdf

Gould, L. (2020). Editorial foreword – Revisiting 'values' in evaluation in times of crisis. *Evaluation Journal of Australasia*, *20*(2), 61–62. https://doi.org/10.1177%2F1035719X20932697

Hehir, G. (2020, April 23). *The Australian government's economic response to COVID-19*. Australian National Audit Office. https://www.anao.gov.au/work/request/the-australian-government-economic-response-to-covid-19

Independent Evaluation Office/United Nations Development Programme & Organisation for Economic Co-operation and Development/Development

Assistance Committee. (2020, April). *Good practices during COVID-19*. United Nations Development Programme. http://web.undp.org/evaluation/guideline/documents/covid19/IEOOECD_DAC_Joint-Guidance_COVID19.pdf

Joint Committee of Public Accounts and Audit. (2020, September). *Annual report 2019–20* (Report 480.1). Parliament of Australia. https://parlinfo.aph.gov.au/parlInfo/download/committees/reportjnt/024492/toc_pdf/Report480.1AnnualReport2019-20.pdf;fileType=application%2Fpdf

Marsh, I. (2006, August). *Can Senate committees contribute to 'social learning'?* (Papers on Parliament No. 45). Parliament of Australia. https://www.aph.gov.au/About_Parliament/Senate/Powers_practice_n_procedures/~/~/link.aspx?_id=AC113D4628DF418F8B4C4AE5782C235A&_z=z

Morrissey, A., Aslam, K., Langford-Smith, J., Chumak, M., & Tilbrook, P. (2021, August 2). *SafeWA – Application audit* (Report 2: 2021–22). Office of the Auditor General Western Australia. https://audit.wa.gov.au/wp-content/uploads/2021/07/Report_2__SafeWA-Application-Audit.pdf

Morrissey, A., Campbell, C., Dias, A., & Fairhead, J. (2020a, December 9). *COVID-19: Status of WA public testing systems* (Report No. 10: 2020–21). Office of the Auditor General Western Australia. https://audit.wa.gov.au/wp-content/uploads/2020/12/Report-10_COVID-19-Status-of-WA-Public-Testing-Systems.pdf

Morrissey, A., Pope, A., Ritchie, M., Wilkins, R., Travia, B., & Fairhead, J. (2020b, December 21). *COVID-19 relief fund* (Report No. 11: 2020–21). Office of the Auditor General Western Australia https://audit.wa.gov.au/wp-content/uploads/2020/12/Report_11_COVID-19-Relief-Fund.pdf

Moulds, S. (2020, May 14). Keeping watch on covid-19 laws: Are parliamentary committees up to the job? *Inside Story*. https://insidestory.org.au/keeping-watch-on-covid-19-laws/

Office of Parliamentary Counsel. (n.d.). *Federal register of legislation*. Parliament of Australia. https://www.legislation.gov.au/Home

Office of the Auditor General Western Australia. (n.d.-a). *Auditing the state's COVID-19 response*. https://audit.wa.gov.au/auditing-in-wa/auditing-the-states-covid-19-response__trashed/

Office of the Auditor General Western Australia. (n.d.-b). *Audit program*. https://audit.wa.gov.au/auditing-in-wa/audit-program/

Parliamentary Joint Committee on Human Rights. (n.d.). *COVID-19 legislative scrutiny*. Parliament of Australia. https://www.aph.gov.au/Parliamentary_Business/Committees/Joint/Human_Rights/COVID19_Legislative_Scrutiny

Phillips, B., Gray, M., & Biddle, N. (2020, August 29). *COVID-19 jobkeeper and jobseeker impacts on poverty and housing stress under current and alternative economic and policy scenarios*. Australian National University Centre for Social Research and Methods. https://csrm.cass.anu.edu.au/sites/default/files/docs/2020/8/Impact_of_Covid19_JobKeeper_and_Jobeeker_measures_on_Poverty_and_Financial_Stress_FINAL.pdf

Rossingh, B. (2021). Editorial. *Evaluation Journal of Australasia*, *21*(2), 65–68. https://doi.org/10.1177%2F1035719X211014398

Schwartz, R., & Mayne, J. (2005). Assuring the quality of evaluative information: Theory and practice. *Evaluation and Program Planning*, *28*(1), 1–14. https://doi.org/10.1016/j.evalprogplan.2004.10.001

Senate Select Committee on COVID-19. (n.d.-a). *Media releases*. Parliament of Australia. https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Scrutiny_of_Delegated_Legislation/Exemptfromoversight/Media_Releases

Senate Select Committee on COVID-19. (n.d.-b). *Public hearings*. Parliament of Australia. https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/COVID-19/COVID19/Public_Hearings

Senate Select Committee on COVID-19. (n.d.-c). *Senate select committee on COVID-19*. Parliament of Australia. https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/COVID-19/COVID19

Senate Select Committee on COVID-19. (2020a, June 15). *Committee letter to commonwealth departments and agencies*. Parliament of Australia. https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/COVID-19/COVID19/Committee_letter_to_Commonwealth_Departments_and_Agencies

Senate Select Committee on COVID-19. (2020b). *First interim report*. Parliament of Australia. https://parlinfo.aph.gov.au/parlInfo/download/committees/reportsen/024513/toc_pdf/Firstinterimreport.pdf;fileType=application%2Fpdf

Senate Select Committee on COVID-19. (2020c). *Terms of reference*. Parliament of Australia. https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/COVID-19/COVID19/Terms_of_Reference

Senate Select Committee on COVID-19. (2021). *Second interim report: Public interest immunity claims*. Parliament of Australia. https://parlinfo.aph.gov.au/parlInfo/download/committees/reportsen/024607/toc_pdf/Secondinterimreport Publicinterestimmunityclaims.pdf;fileType=application%2Fpdf

Tham, J.-C. (2020, April 26). COVID-19 ministerial powers need to be reined in. *The Canberra Times*. https://www.canberratimes.com.au/story/6732299/covid-19-ministerial-powers-need-to-be-reined-in/?cs=14264

United Nations Office of Internal Oversight Services. (2020, June 12). *Synthesis of guidelines for UN evaluation under COVID-19*. United Nations Evaluation Group. http://www.uneval.org/document/download/3620

Wilkins, P., Gilchrist, D., & Phillimore, J. (2020). Independent review of emergency economic stimulus measures: Global financial crisis and COVID-19. *Australian Journal of Public Administration*, *80*(1), 12–28. https://doi.org/10.1111/1467-8500.12437

9 Evaluation for Systems Transformations

Lessons from the Pandemic

Michael Quinn Patton

Past lessons ignored

By the end of 2021, over 300 million confirmed COVID-19 cases and over 5 million deaths have been reported since the start of the pandemic with no end in sight. The talk and hope, ironically, is of the virus becoming "merely" endemic like the seasonal flu rather than a raging pandemic. The COVID-19 pandemic has provided a glimpse into the magnitude of changes set in motion by a global emergency. United Nations Secretary-General Antonio Guterres (United Nations, 2021), among many others, has warned consistently throughout the pandemic that climate change looms over the world as a larger, more far-reaching global emergency for which COVID-19 has been but a dress rehearsal, an early warning of what lies ahead at greater magnitude though slower manifestation.

From the outbreak of the pandemic, the search for previous lessons was underway (Chelsky & Kelly, 2020; Independent Evaluation Group [IEG], 2020), including how little had been learned from the Spanish flu pandemic in 1918. Much was learned about mitigating the global HIV/AIDS epidemic, lessons relevant to addressing the COVID-19 crisis despite the different nature of HIV transmission, but those lessons were largely ignored (Rugg et al., 1999). In essence, fundamental prevention and mitigation principles flowing from epidemiology and evaluation still apply, ignored though they may be by contemporary politicians (Mukherjee, 2020). For example, many prior lessons had been captured and thoughtfully articulated as in the *Field Epidemiology Manual* of the Centers for Disease Control and Prevention (CDC) (Rasmussen & Goodman, 2018). The *Manual* devotes an entire chapter to communication during a health emergency, asserting that there should be a lead spokesperson whom the public gets to know – familiarity breeding trust. The spokesperson should have a "Single Overriding Health Communication Objective," or SOHCO (pronounced sock-O), which should be repeated at the beginning and the end of any communication with the public. After the opening SOHCO, the spokesperson should "acknowledge concerns and express understanding of how those affected by the illnesses or injuries are

DOI: 10.4324/9781003376316-10

probably feeling." Such a gesture of empathy establishes common ground with scared and dubious citizens who, because of their mistrust, can be at the highest risk for transmission. The spokesperson should make special efforts to explain both what is known and what is unknown. Transparency is essential, the field manual says, and officials must "not over-reassure or overpromise."

Yet in the United States, President Donald Trump and other political appointees used pandemic press conferences for political messaging and grandstanding, often undercutting or contradicting the scientific explanations and advice of CDC epidemiologists. As the pandemic grew in severity, the politicians turned to blaming and attacking the scientists at CDC. The result has been increased political polarization, resistance to basic public health measures (e.g., social distancing, mask-wearing), and a dangerous anti-vaccine movement.

This post-truth, anti-science world

The pandemic has accelerated and made transparent the dangerous ramifications of a post-truth, anti-science world. Information, knowledge, and science have all been politicized at a societal level. This "post-truth world" is characterized by the politicization, corruption, and suppression of science. Scientific and evaluative evidence are treated as no more valid than personal opinion. Political advocates promulgate their own "facts." The distinction between evidence and opinion has become politically blurred. We face an "infodemic" of inaccurate, misleading, unsubstantiated, and dangerous medical advice on social media opposing masking and vaccinations. The infodemic is fed by fake news, distortions of facts, outright lies, and outrageous conspiracy theories. Misinformation breeds distrust and suspicion and undermines evidence-based decision-making. Distortions of reality, denial of facts, and attacks on scientific truth fuel racism, misogyny, homophobia, classism, and polarization. Misinformation, bad data, distorted statistics, and well-meaning but wrong interpretations are rampant, dangerous, and, in a health emergency, can cost lives.

Trustworthy, valid, and useful information is at a premium in a crisis. We all, as evaluators, bear societal responsibility to serve the public good. This means staying informed, being a fact checker, and helping ensure that *facts trump ideology and politics.* This goes beyond conducting any singular evaluation to our collective responsibility to society as evaluation scientists. After a speech I gave as COVID was first emerging, the sound technician came up to me and asked if I knew that the virus came from a Chinese secret chemical weapons facility built on an ancient civilization where they were mining vicious microbes. I invited him to sit with me for a minute and took him to an internet site that debunked that story to his eventual satisfaction.

That is the macro view of politics and evaluation – and as evaluation professionals we are all engaged in this larger societal battle about what

constitutes evidence and truth in an increasing post-truth, anti-science political world. That is the context we should consider as we extract lessons about transforming evaluation to deal with these challenges.

Evaluating systems transformation requires transforming evaluation

Given the pandemic's global impact, the looming climate emergency and related threats to a just and sustainable world, systems transformation has become the clarion call of our times.

The transformation represented by global warming can be captured in photographs of melting glaciers, satellite images of barren, brown land once covered in ice, and graphs of annual temperature increases. The transformation of plastic into ocean pollution is represented by the Great Pacific Ocean Garbage Patch, 80,000 tons of discarded plastic covering an area of about 617,800 square miles (1.6 million square kilometers), a vortex of micro-particles swirling in a gyre of marine debris. The process of transforming the lush Amazon rain forest, the Earth's biodiversity-rich lung, into a massive despoiled and degraded landscape is visible in large fires and denuded land. The Anthropocene is a geologic epoch defined by the theory that the human species has transformed the ecosystem function of our planet in ways that are unsustainable. That transformation has accelerated dramatically in just one human generation – ours! Now we need transformation to reverse the negative human effects on the planet.

The theme of the 2019 conference of the International Development Evaluation Association (IDEAS, 2019) was "Evaluation for Transformational Change" and it generated a book with the same title (Berg et al., 2019). A subsequent IDEAS book was entitled *Transformational Evaluation for the Global Crises of Our Times* (Berg et al., 2021). Evaluators enter the fray to assess the fidelity and impacts of hypothesized transformational initiatives and trajectories. But transformational initiatives offer new challenges for the design, implementation, and use of evaluations. The premise of this chapter is that *evaluating transformation requires transforming evaluation*. I will offer three overarching evaluation transformations that are needed: moving from project thinking to systems thinking, moving from theory of change to theory of transformation, and engaging seriously with the implications for evaluation of complexity.

Three overarching evaluation transformations

From project thinking to systems thinking

Transformation is not a project or program. Transformational initiatives are not targeted to achieving SMART goals that are specific, measurable, achievable, realistic, and time bound. Transformation means changing

systems to be more just (equitable) and sustainable (resilient) (Patton, 2020a, 2021a, 2021b). This means dealing with complex dynamics in a world characterized by turbulence, uncertainty, unpredictability, and uncontrollability (Furubo et al., 2013; Hodgson, 2020). The focus of evaluation, the *evaluand* in evaluative practice, becomes transformed systems.

The COVID-19 pandemic has increased the flow of private sector funds into systems transformations (Olazabal, 2021; The Investment Integration Project, 2020). We're seeing emphasis on systems change wherever serious actors are addressing the climate emergency. For example, the global financial investment community has been highlighting changes in their sphere as discussed in *Assessing System-Level Investments* (Lydenberg & Burckart, 2020). As the report shows, trillions of dollars are being directed at systems level change and social impact investors are seeking new approaches to assess such changes.

The Systems in Evaluation Topical Interest Group of the American Evaluation Association spent two years identifying the principles that constitute systems thinking: focusing on interrelationships, perspectives, boundaries, and dynamics (Systems in Evaluation, 2018).

Among many other things, the global pandemic powerfully demonstrated the interconnections among health systems, school systems, community systems, economic and finance systems, entertainment systems, and political systems. At any given moment, the focus tended to be on some discrete and particular solution like wearing masks, social distancing, more testing, quarantining the sick, and flattening the curve. However, the entire health system was in crisis, an emergency that emerged and rapidly accelerated from years of neglect, ignored warnings, and under-resourced existing health systems at all levels (Tufekci, 2021). The pandemic epitomizes what it means to operate scientifically and evaluatively in a complex dynamic systems emergency. Consider the nature of epidemiology and what evaluators can learn from that esteemed and crucial profession:

> Epidemiology is a science of possibilities and persuasion, not of certainties or hard proof. "Being approximately right most of the time is better than being precisely right occasionally," the Scottish epidemiologist John Cowden wrote, in 2010. "You can only be sure when to act in retrospect" ….
>
> Epidemiologists must persuade people to upend their lives—to forgo travel and socializing, to submit themselves to blood draws and immunization shots—even when there's scant evidence that they're directly at risk ….
>
> Epidemiologists also must learn how to maintain their persuasiveness even as their advice shifts. The recommendations that public-health professionals make at the beginning of an emergency—there's no need

> to wear masks; children can't become seriously ill—often change as hypotheses are disproved, new experiments occur, and a virus mutates.
> (Duhigg, 2020, p. 17)

Evaluators have much to learn from epidemiologists about how to engage in complex, dynamic systems during emergencies, which is the world we are likely to all face with the exacerbating global climate emergency going forward. Public health, community health, national health, global health, family health, and personal health are all connected. This is micro to macro, and macro to micro, systems thinking. The state of public health is also connected to the economy, the financial system, politics at every level, social well-being, cultural perspectives, educational inequities, social and economic disparities, public policy decisions, and evaluation.

Using a systems lens: The example of the UN Food Systems Summit

A critical evaluation skill is being able to see the interconnections among systems and the implications of those interconnections. The United Nations held a global Food Systems Summit on September 23, 2021. Building up to the Summit, more than 900 "Independent Dialogues" were held around the world. The synthesis of those Dialogues confirmed the importance of taking a systems perspective and seeing the interconnections of food systems with health, climate, social justice, and information systems. There are nearly 690 million people in the world who are hungry, or 8.9 percent of the global population – an increase of 10 million people in one year and nearly 60 million in five years – and the COVID-19 pandemic has only exacerbated the problem. Food systems transformation involves changing systems. The importance of thinking in terms of systems permeated the Dialogues. The pandemic gave rise to many discussions of the interconnections between health systems and food systems, including the significant increases in food insecurity and hunger due to COVID.

The Food Systems Summit elevated and focused attention on food systems, not just food. The very framing of the Summit, and therefore the framing of the Independent Dialogues, drew attention to food systems, not just food production and consumption. As a result, the language of systems permeated the Dialogues. The Dialogues were occurring during the COVID pandemic and increased evidence of the climate emergency with increased incidence of severe weather episodes, fires, droughts, and floods. Progress on the Sustainable Development Goal (SDG) indicators was often reversed, as great numbers of people experienced food insecurity, hunger, and deepened poverty. Dialogue participants often observed that the potential for food systems transformation was inevitably and intrinsically tied to the transformation of climate and health systems. Dialogues addressed a broad range of the needed systems transformation

from national-level systems to community-level systems, including marketing systems, seed bank systems, land tenure systems, and finance systems (Patton et al., 2021).

However, while the Food Systems Summit elevated and focused attention on food systems and the language and rhetoric of systems was noticeably in the ascendant, the report observed that *thinking* in systems was noticeably absent. The transition from simple, linear project and program thinking to systems thinking constitutes a substantial change in worldview. It is a paradigm shift of major proportions (Patton et al., 2021).

Systems thinking means designing, implementing, and evaluating transformation initiatives with attention to the interdependencies among humans and nature, and among producers, distributors, and consumers of food. Systems thinking maps and incorporates diverse perspectives within and across ecosystem, political, economic, social, and cultural boundaries. It identifies and monitors the dynamic interactions of multiple factors and relationships in the production and consumption of food, attending to iterative interconnections, feedback loops, leverage points, momentum dynamics, critical mass transitions, networked interactions both formal and informal, and cross-silo interconnections among multiple stakeholder constituencies: governments, private sector actors, civil society and non-government organizations (NGO), advocates and activists, researchers and university scholars, philanthropic donors and social impact investors, international and domestic agencies involved in various aspects of food systems, and managers and evaluators of transformational initiatives. Applying systems thinking requires understanding and acting on the interdependent nature of land, air, and water systems and the knowledge that food systems transformation is connected to climate change, health systems, sustainable ecosystems, weather systems, and healthy landscapes and seascapes. Transforming complex systems interconnections requires a theory of transformation, the second overarching evaluation transformation.

From theory of change to theory of transformation

A theory of change specifies how a project or program attains desired outcomes. Transformation is not a project. It is multi-dimensional, multifaceted, and multi-level, cutting across national borders and intervention silos, across sectors and specialized interests, connecting local and global, and sustaining across time. A theory of transformation incorporates and integrates multiple theories of change operating at many levels that, knitted together, explain how major systems transformation occurs.

Program theory aims to explain why a particular program approach should work to achieve desired results. This involves making explicit and then testing a program's theory of change. In 1995, Carol H. Weiss, an applied sociologist and pioneering evaluation theorist who helped create the field of evaluation, wrote an article for the Aspen Institute about the

importance of basing community interventions on a solid theory of change. Her article was entitled: "Nothing as practical as a good theory" (Weiss & Connell, 1995). She was reacting to the emergence of large-scale community initiatives funded by philanthropic foundations and government agencies that poured millions of dollars into community change efforts with no knowledge of the relevant social science research that should have been informing such efforts. Her article became one of the most influential, if not the most influential, article in the history of program evaluation.

But transformation involves a different order of magnitude and speed than project-bounded changes – and, correspondingly, requires a different kind of theory. The language of transformation suggests major systems change and rapid reform at a global level. A transformational trajectory would cut across nation states, across SDG and sector silos, and connect the local with the global (using the Blue Marble principles of evaluation discussed in my book on the subject) (Patton, 2020a). The language of transformation has emerged across the globe wherever people convene to contemplate and initiate collective action to deal with global issues. A vision of transformation has become central to international dialogues about the future of the Earth and sustainable development.

A theory of transformation emerges from studying major transformations of the past and examining current challenges and patterns that portend future possibilities. Transformations that are instructive include the end of colonialism, the end of apartheid, the fall of the Berlin Wall and communism, turning back the AIDS epidemic, the internet and, today, social media. It is instructive to understand how these systems emerged into dominance in the first place, for none of these transformations occurred due to a centrally conceptualized, controlled, and implemented strategic plan or massive coordinated initiative. These transformations occurred when multiple and diverse initiatives intersected and synergized to create momentum, critical mass, and ultimately tipping points.

New kinds of initiatives and new forms of intervention will be needed that can respond to the challenges of global problems, including designing and evaluating systems transformations. Transformation flows from an understanding that the status quo is not a viable path forward and that networked action on multiple fronts using diverse change strategies across multiple landscapes will be needed to overcome the resistance from those who benefit from the status quo. Multiple interventions are required to multiply effects, creating streams of diverse interventions flowing together to generate critical mass tipping points and mammoth change in global systems. Thus, transformation is simultaneously and interactively global and local at the same time, and contextually sensitive and rooted while being globally manifest and sustainable. Transforming systems must be multi-faceted, multi-dimensional, multi-sectoral, multinational, and multiplicative. Tracking these new, transformational initiatives will require complex global systems change approach to evaluation.

Transformation is a sensitizing concept that must be given meaning and specificity within the context where transformation is targeted. Evaluation of transformation begins by examining whether an initiative, or more likely a set of initiatives and interventions, constitute *a trajectory toward transformation*. Asking the trajectory question changes the evaluation focus from transformation having occurred (or not) to *transformational engagement*. That is the reframing formulated by the influential Independent Evaluation Group (IEG) of the World Bank. Assessing the trajectory toward transformation is what most funders, decision-makers, and implementers of initiatives are looking for from evaluation.

> Transformational engagement is an intervention or a series of interventions that helps achieve deep, systemic, and sustainable change with large-scale impact in an area of a major development challenge. These engagements help clients remove critical constraints to development; cause or support fundamental change in a system; have large-scale national or global impact; and are economically, financially, and environmentally sustainable.
>
> (IEG, 2016, p. 1)

The IEG evaluated a sample of 20 transformational engagements varying in form, size, the development challenges they address, sector, and region, as well as country context. In addition, IEG reviewed a purposeful and selective sample of country-level engagements. Their comparative and synthesizing analysis exemplifies systems transformation evaluation (IEG, 2018; see also Heider, 2017; IEG, 2016).

The Global Alliance for the Future of Food (the "Global Alliance") formulated a theory of transformation aimed at stimulating and integrating local and global food systems transformations. The Global Alliance is formed of 30 philanthropic foundations that collaborate to support the transformation of food and agricultural systems. "Transformational means realizing healthy, equitable, renewable, resilient, and culturally diverse food systems shared by people, communities, and their institutions" (Richardson, 2019). In January 2020, the Global Alliance formally adopted a theory of transformation that informs its activities and provides a basis for evaluating its products, activities, and impacts through the lens of transformational engagement (Global Alliance for the Future of Food, 2020).

The synthesis of Independent Dialogues generated by the UN Food Systems Summit also generated a theory of transformation that integrates 22 guiding themes that together hypothesize how to mobilize and accelerate food and agricultural systems transformation (Patton et al., 2021).

It is beyond the scope of this chapter to present a theory of transformation. I would simply reiterate that a theory of transformation synthesizes multiple theories of change. Any specific theory of change concerns how to produce desired results targeted by a particular intervention.

Transforming systems requires aligning, networking, and integrating multiple and diverse theories of change to build critical mass transformational tipping points. Transformation, then, is not an intervention; it is rather a movement creating synergies among multiple interventions (Patton, 2020a).

Inform and infuse evaluation with complexity understandings

The two evaluation transformations discussed above – moving from project thinking to systems thinking and moving from theory of change to theory of transformation – are grounded in complexity understandings. Evaluation is dominated by linear causal modeling and thinking. The action paradigm is one of control: plan your work, work your plan. Complexity theory involves and addresses nonlinearities, emergence, and lack of control (inherent dynamic complex system uncertainties). Understanding any and all aspects of the pandemic requires complexity theory understandings and insights. Evaluation under conditions of complexity is different from traditional linear static models of interventions and evaluation (Bamberger et al., 2016; Patton, 2011, 2020a).

As the pandemic emerged, its scope, spread, speed, and endurance were unknown. Social distancing affected all aspects of human interactions at the personal, family, organizational, community, and international levels. "Pivoting" gained momentum as a response strategy with the spread of the virus affecting employment, access to services, schools, businesses, planned public events of all kinds, and public transportation, to name but a few arenas of impact. Lockdowns, quarantines, mask mandates, and, later, vaccine mandates affected all aspects of society. Different countries responded in different ways. Hospitals were quickly overwhelmed. Poverty levels worldwide increased dramatically for the first time in a decade. Millions were thrown into food insecurity.

Meanwhile, the coronavirus evolved into the more virulent Delta variant which created a new wave of infections and increased deaths. The emergence of the vaccines manifests major inequities as richer countries benefited quickly and poor countries had limited access. Health disparities between socio-economic, racial, ethnic, and immigrant groups were accentuated. Governments, NGOs, humanitarian agencies, businesses, philanthropic foundations, universities, and communities scrambled to respond to the deepening pandemic. Diverse actions and interventions led to both intended and unintended consequences, and ripple effects that fed on each other and exacerbated the devastating effects of the virus. Complexity reigned – and it still does.

In March 2020, I wrote in a blog about the implications of the pandemic for evaluation from a complexity perspective. I noted that evaluators would have to be prepared to pivot, adapting evaluation plans and designs, and become capable of responding to complex dynamic systems. This means

being prepared for the unknown, for uncertainties, turbulence, lack of control, nonlinearities, and for emergence of the unexpected. This is the current context around the world in general and this is the world in which evaluation will exist for the foreseeable future (Patton, 2020b). This means agility rules. Here are principles I propose to inform and undergird evaluation agility.

Five principles for evaluation agility

Timely data rules

Channel a sense of urgency into thinking pragmatically and creatively about what data can be gathered quickly and provided to evaluation users to help them know what's happening, what's emerging, how needs are changing, and consider options going forward. At the same time help them document the changes in implementation they are making as a result of the crisis – and the implications and results of those changes. You may be able to gather data and provide feedback about perceptions of the crisis and its implications, finding out how much those affected are on the same page in terms of message and response. That's what developmental evaluators do.

Be adaptable

Expect change. Program goals may appropriately change. Measures of effectiveness, target populations, implementation protocols, and outcome measures all may change. This means that evaluation designs, data collection, reporting timelines, and criteria will and should change. Intended uses and even intended users may change. Expect and facilitate change, document changes and their implications. That is an evaluator's job in a crisis, not to continue a comfortable, business-as-usual mindset; there is no business-as-usual now. If you don't see program adaptation, consider pushing for it by presenting options and introducing scenario thinking at a program level. Take risks, as appropriate, in dealing with and helping others deal with what's unfolding.

Think globally, act locally

Context matters for every evaluation, but what is involved in contextual assessment has now expanded to a global level. Global patterns and trends intersect with and affect what happens locally, including evaluations at whatever level one is operating. Zoom out to understand the big picture of what's happening globally and zoom in to the implications locally, where locally means whatever level one is working at. The whole world is now part of the evaluation context. The Global South and Global North will be intertwined as never before as the global health emergency deepens and broadens.

Prepare to make the case for evaluation's value

Expect proposals to cut back evaluation funding. One of the first targets for budget cuts in recessions and political turmoil has historically been evaluations. Prepare to make evaluation all the more useful so that *the evaluation value proposition* reframes evaluation as an essential activity, not as a mundane bureaucratic or luxurious function when times are good. Define, conceptualize, articulate, and demonstrate the essential utility of evaluation when judgments are premature, and when the facts are uncertain. Refrain from expressing uninformed or premature judgments and urge others to do likewise.

Advocate for better data

Reports of the incidence and prevalence of COVID appeared to be problematic in many cases. Ongoing systematic, stratified random sample testing will be needed to establish population infection rates. Understand the strengths and weaknesses of epidemiological statistics as well as other indicators relevant to any particular crisis.

An ethical framework for transformational evaluation

Moving from project thinking to systems thinking, from theory of change to theory of transformation, and from simple linear causality to complex dynamic systems understandings provides an analytical and conceptual framework for evaluating transformation and transforming evaluation. What remains is to ensure that the transformational engagement is ethically grounded and appropriate. For evaluators there are two sources that express shared professional and ethical values: (1) global commitments and values manifest in the global Agenda 2030, SDGs, and in treaties and declarations protecting the rights of marginalized populations and (2) the standards and principles adopted by the evaluation profession. Those two sources together have made *equity* and *sustainability* the cornerstones of the global common good. In the new 5th edition of *Utilization-Focused Evaluation*, all evaluators are called on to address equity and sustainability as universal evaluation criteria (Campbell-Patton & Patton, 2022, ch. 18). Each criterion can be reviewed through an ethical lens.

Equity

Calls for transformation flow from two streams: one values-based and visionary, the other crisis-focused and fear-of-calamity driven. Transformation as a values-based vision flows from the hopes expressed in the *Universal Declaration of Human Rights* (adopted in 1948) and subsequently in the United Nations *Declaration of the Rights of the Child* (adopted

in 1959). Global diversity, equity, and inclusion (*DEI*) norms and values are expressed and codified in the United Nations *Declaration on the Rights of Indigenous Peoples* and the *International Women's Bill of Rights*. All people – young and old – have the right to food, water, sanitation, security, shelter, respect, and dignity. As expressed in the ambitious SDGs, adopted in 2015, entitled *Transforming Our World* (United Nations [UN], 2015), transformation means leaving no one behind (Segone & Tateossian, 2017).

Caroline Heider, as former Director General Evaluation at the World Bank Group, has considered this criterion and its implications in depth:

> Although the [OECD-DAC] evaluation criteria appear to be neutral and should be applied as such, they were informed by a set of values. The post-2015 agenda has declared its intention to be more inclusive, respecting underprivileged groups of people, which means we as evaluators need to reflect whether the criteria suit these intentions. Being able to shape norms that are more inclusive of diversity rather than judge everyone through more limiting norms will be a necessity if 2030 is to become the world we want.
>
> (Heider, 2017, p. 5)

United Nations Children's Fund (UNICEF) and other international agencies have promoted equity-focused evaluation based on human rights and the rights of children (Bamberger & Segone, 2011). This vision for evaluation's role in the world was articulated in the theme of the 2014 annual conference of the American Evaluation Association: Visionary Evaluation for a Sustainable, Equitable Future. Two important evaluation thought leaders, Stewart Donaldson and Robert Picciotto (2016) edited a book on *Evaluation for an Equitable Society*.

The Equitable Evaluation Initiative (2021) promotes the use of evaluation as a tool for advancing equity. Equitable evaluation encourages evaluators to consider four aspects in their evaluation practice, all at once: diversity of evaluation teams (beyond ethnic and cultural); cultural appropriateness and validity of evaluation methods; ability of evaluation designs to reveal structural and systems-level drivers of inequity; and the degree to which those affected by what is being evaluated have the power to shape and own how evaluation happens (Dean-Coffey, 2018; Equitable Evaluation Initiative, 2021).

The *DEI* criterion can include any or all of several such important perspectives:

- Mertens' (1999, 2009) *transformative evaluation paradigm* aimed at ensuring equity for the diverse voices of people historically marginalized.
- Dealing with racism and white privilege, including white frames in evaluation language (Johnson, 2019; S. Shanker, 2019; V. Shanker, 2019).
- *Culturally responsive evaluation* (Hood et al., 2015).

- The NICE framework – *Navigating the Intersection of Culture and Evaluation* – which addresses culture at national or transnational levels to consider "societal or national dispositions rather than one single culture" (Ofir, 2018c, p. 1).
- *DEI* concerns inclusion of diverse people and perspectives from the Global South in pursuit of global equity, an example is Made in Africa Evaluation (Ofir, 2018b).
- *Decolonizing evaluation*, included by *DEI* through both development initiatives and, correspondingly decolonizing evaluation (Chouinard & Hopson, 2016; McKegg, 2019). Decolonizing methodologies (Smith, 1999) aim to redress inequities and misrepresentations manifest in research on indigenous peoples and evaluation of programs targeted at indigenous populations.
- The United Nations Evaluation Group Norms and Standards incorporate a rights-based criterion for evaluation:

> *Norm 8: Human rights and gender equality.* The universally recognized values and principles of human rights and gender equality need to be integrated into all stages of an evaluation. It is the responsibility of evaluators and evaluation managers to ensure that these values are respected, addressed and promoted, underpinning the commitment to the principle of "leave no one behind."
>
> (UN Evaluation Group, 2016, Norm 8)

Developed by Khalil Bitar (2021), a leader of EvalYouth, the *Social Equity Assessment Tool (SEAT) for Evaluation* consists of 13 equity dimensions that assess:

> The equitable treatment of relevant community members and right-holders/right holder groups within the broad geographical area the intervention covers and meaningfully involving them in the intervention design and implementation. The SEAT consists of eight demographic aspects (geographical, economic, gender, racial and ethnic, religion, age, sexual orientation, and disability) and five cross-cutting aspects (intervention team, evaluator/evaluation team, data collection/analysis/reporting, environmental justice, and unintended consequences).
>
> (pp. 6–7)

This tool and framework for applying and assessing equity criteria is meant to be used universally whether the intervention has explicit equity goals or not. Indeed, Bitar asserts, "it is even more necessary to use a SEAT when the intervention does not have equity-related objectives" (p. 7). Doing so ensures that attention to equity is a universal evaluation criterion.

Sustainability

As noted earlier, the theme of the 2019 conference of IDEAS in Prague was *Evaluation for Transformative Change*. At the conclusion of the conference, participants from around the world adopted a "Declaration on Evaluation for Transformational Change." The Declaration was adopted on October 4, 2019, in Prague and included a commitment to address sustainability in all evaluations:

> In all our evaluations, we commit to evaluating for social, environmental and economic sustainability and transformation, including by assessing contextual factors and systemic changes. We commit to assessing and highlighting, in all evaluations, unintended negative social, economic and environmental effects.
>
> (Item 6 of 10 in the Declaration. For the full declaration, see International Development Evaluation Association, 2019)

All evaluations, with an emphasis on *all*, are mandated to include attention to sustainability, specifically ecosystem sustainability. The global climate emergency, according to the IDEAS Declaration, requires action and engagement by everyone, everywhere, including evaluators.

Adaptive ecological sustainability has emerged as a priority criterion for evaluation (Ofir, 2017, 2018a; Rowe, 2019; Uitto, 2019). A volume of *New Directions for Evaluation* is centered on sustainability (Julnes, 2019).

Interdependence of equity and sustainability

Equity and sustainability are not competing criteria. They intersect, overlap, and are mutually reinforcing. Sustainability and equity, combined, are the foundation for transformation. This relationship links sustainability to equity and transformation. For example, Amnesty International established tackling the climate crisis by supporting a "human rights-centered transition to a green economy" as its top priority for 2020 (Dubb, 2020).

Evaluation as a profession suffers its own history of racism and white supremacy. Going "blue" (Blue Marble Evaluation) and "green" (environmental sustainability as a criterion) does not exempt us from dealing with Black, Brown, and White; it is quite the contrary. To decolonize evaluation (Chouinard & Hopson, 2016), culturally responsive and equitable evaluation must be part of an evaluation commitment to and engagement with sustainability for human survival on Earth. So, concern for sustainability of the Earth and humanity is connected to diversity, equity, and inclusion.

Evidence abounds that the most marginalized and vulnerable populations are those who are and will continue to be most negatively affected by climate change. UNICEF Executive Director Anthony Lake introduced a report on the impact of climate change on children, entitled *Unless We*

Act Now, with an overview of the threat to children experiencing poverty worldwide:

> In every crisis, children are the most vulnerable. Climate change is no exception. As escalating droughts and flooding degrade food production, children will bear the greatest burden of hunger and malnutrition. As temperatures increase, together with water scarcity and air pollution, children will feel the deadliest impact of water-borne diseases and dangerous respiratory conditions. As more extreme weather events expand the number of emergencies and humanitarian crises, children will pay the highest price. As the world experiences a steady rise in climate-driven migration, children's lives and futures will be the most disrupted.
>
> (UNICEF, 2015, p. 6)

The Independent Dialogues as part of the UN Food Systems Summit also addressed issues of equity and sustainability. The top three themes in the synthesis of the Dialogues were as follows:

1. *Transform food systems* – Transformation meaning major, significant, deep, and broad changes beyond piecemeal reforms, incremental change, and narrowly focused projects and programs.
2. *Ensure sustainability and strengthen resilience* – Sustainability meaning humanity and nature thriving together. Resilience meant capacity to regenerate and adapt. Resilience supports sustainability.
3. *Make equity a priority* – Dialogue participants emphasized contributions to equity as a priority criterion for judging food systems solutions.

Guarantee the right to food

In support of these themes, most Dialogues focused discussions on pathways for making food systems more sustainable and equitable through transformative solutions in production and consumption, through changed policies and innovations, and by engaging in multi-stakeholder collaborations. Dialogues that discussed the right to food began with the premise that the framework for transformation already exists in the *Universal Declaration of Human Rights*. Conceptualizing food as a right rather than merely a market-based commodity would provide a unified and universal framework for food systems transformation.

Utilization-focused evaluation

The implication for utilization-focused evaluation is that evaluators present to primary intended users the emergence and importance of equity and sustainability as evaluation criteria and facilitate discussion of how

those criteria can be addressed in the evaluation by making a commitment to promoting each of them as intended uses of the evaluation. *Utilization-focused evaluation* going forward commits to making contributions to equity and sustainability as criterion of both program and evaluation excellence and success (Campbell-Patton & Patton, 2022).

Utilization-focused evaluation is driven by the obligation and opportunity to meet the information needs of primary intended users to enhance use of evaluation and extend evaluative thinking. Now, however, in the face of the pandemic; climate emergency; global social justice uprising; worldwide food systems failures; and the dramatically increasing gap between the haves and have-nots, means the active-reactive-interactive-adaptive framework of utilization-focused evaluation includes addressing the criteria of equity and sustainability. Facilitation must undergo what amounts to a paradigm shift. Evaluators are not just responsible for meeting the information needs of primary intended users. We now have the additional obligation to bring before primary intended users the larger societal issues of sustainability and equity. This obligation flows from adoption by evaluation professional organizations of updated statements on our professional responsibilities because of what's at stake for humanity, not just for primary intended users. Here's an example of engaging these universal criteria at a local level.

In my home in Minnesota, I often open training and engagement sessions with primary intended users by asking about fishing. Minnesota's landscape is dominated by thousands of lakes created when the glaciers of the last Ice Age receded. The state motto is the *Land of 10,000 Lakes.* Fishing is the most common outdoor activity in Minnesota; some 80 percent of the population goes fishing at least once a year. I ask about fishing by positing that every issue in evaluation can be illustrated and illuminated by fishing.

Success criteria: What constitutes a good day's fishing? Number of fish caught? Size of fish caught? Type of fish caught?
Alternative outcomes: Fishing for food versus fishing for recreation (catch and release).
Process considerations: Who do you fish with? Family? Friends? Children?
Implementation variations: Fishing from shore; fishing from a boat; ice fishing in winter.
Stakes: Food; personal satisfaction; healthy lifestyle; participating in fishing competitions with large cash prizes.
Cost-benefit issues: People can spend thousands of dollars on equipment to catch fish that cost $15 a pound in the supermarket. How do people fishing calculate the cost and benefits?

This opening exercise introduces evaluation issues and evaluative thinking through an activity with which a great many people have had direct

experience, and even those who don't fish understand the relevance. That version of the exercise simply focuses on standard evaluation logic. Now, however, to introduce the criterion of sustainability, I have reframed the exercise beyond asking about how fishing is going. I ask also: what are you observing about the health of the fishing ecosystem? Minnesota's' state fish is the walleye, a prized fresh-water game fish. The walleye is sensitive to temperature. An increase in lake temperature due to climate change threatens the walleye population. Other trends affecting fishing ecosystem include runoff into lakes of chemicals used in agricultural production; more severe weather with heavy rains and fierce winds affect lake water quality; increased pollution of lakes, not only from sources within Minnesota, but chemicals like mercury carried in the atmosphere by prevailing global winds; plastic microfibers found in lake water and fish; invasive species like zebra mussels that threaten fish populations; and loss of habitat. Through farming, transportation, coal-burning electrical plans, and heating and cooling buildings, the Midwestern part of the United States, which includes Minnesota, emits more greenhouse gases into the atmosphere than the population-dense Northeast, the oil-field states of Texas and Oklahoma, the wildfire-prone West (California, Oregon, Washington state), and the big air conditioner users in the hot Southwest (Sturdevant, 2021). This comparison comes as a surprise, even a shock, to most Minnesotans.

This becomes a way to introduce issues of sustainability and to address how a particular project or program will have implications for the climate emergency. Likewise, the obligation to address equity can be drawn out of fishing. The most valuable property in Minnesota is lakefront property. Who owns lakefront property? Who has access to lakes for recreation and fishing? Treaty issues with Native Americans about fishing rights are a long-standing area of dispute with incursions by the surrounding non-Native American populations onto reservation territories to fish and hunt. Public funds spent on protecting lakes support whom? What's the relationship between property values and lake access? Mercury in fish poses a particular threat to immigrant populations that may rely heavily upon fishing for food.

The point here is for evaluators to use their own knowledge of the context within which a particular evaluation is taking place to facilitate consideration of whether and how attention to the criteria of equity and sustainability can be included in the evaluation, any evaluation.

Evaluating transformation requires solid ethical grounding (Patton, 2022). The ethics of transformation involve interconnections between personal ethics (transforming our own behaviors), professional ethics (actively advocating a transformational stance among professional evaluators), societal ethics (examining evaluation's role in support of the public good and democratic processes), and global ethics (ensuring attention to and engagement with the global emergency by incorporating transformational

criteria of equity and sustainability into all evaluations based on human rights). This chapter has examined the implications of transformative ethics for evaluation theory, practice, and methods.

Evaluation and evaluators as part of the transformation process

Having "skin in the game" means having a personal stake in the outcome; it means you are a stakeholder. When it comes to the survival of humanity and the planet, we all have skin in the game as we and our loved ones live in the world that is under threat. We are not outside looking in. We are part of the global system and there's a good chance that we are each, in our own way, part of the problem. This gives us a quite different stance than is typically expected. Evaluators are virtually always outside the programs or projects they evaluate. Acknowledging and facing the realities of the need for major systems changes transforms the position of evaluators from external observers of change to internal participants in change.

Traditionally the evaluator's credibility flows from independence and neutrality. Evaluation for transformation changes the evaluator's role and credibility, based on interdependence, and being involved. There is no external, independent stance in a pandemic. Everyone is affected, everyone has a stake, including evaluators.

We are facing immense global challenges rooted in the legacies of colonialism and white supremacy. Extractive and exploitative practices have led to deep inequalities based on race, geography, class, gender, and many more divisions, and a rapidly changing climate that threatens biodiversity and humanity itself. What, then, is the role of evaluation in addressing these challenges? It begins with a recognition that evaluation is not (and has never been) value neutral.

In 2004, eminent evaluation scholar Robert Stake published a provocative article that asked: "How Far Dare an Evaluator Go Toward Saving the World?" (Stake, 2004). His question raised the issue of what role evaluators' values play in the design and conduct of evaluations. Facilitating clarification of intended users' values as a foundation for designing and enhancing use of evaluations is a central feature of utilization-focused evaluation. A second dimension of valuing concerns what role evaluators' values play. A third-dimension concerns how values adopted by the evaluation profession are brought into the design of evaluations as discussed earlier. We all have a stake in a more just and sustainable world. Stake asked how far an evaluator dare go toward saving the world. A broadening of that question in the context of our current pandemic and climate emergencies becomes: how far dare we, collectively, as an evaluation profession go in changing the world? Are we prepared to transform evaluation to play our role in evaluating transformation? (Patton, 2021b).

Conclusion

The premise of this chapter is that *evaluating transformation requires transforming evaluation*. The greatest danger for evaluators in times of turbulence like the pandemic is not the turbulence itself – it is to act within yesterday's paradigm without adapting evaluation to the challenges of a changed world. Transformational initiatives offer new challenges for the design, implementation, and use of evaluations. The nature of the transformations that emerge will be mediated by context. The evaluation architecture which determines the demand, supply, and nature of evaluative products is quite varied. But the need for transformation at some level, in some ways, to meet the challenges of creating a more just and sustainable world is universal.

The ultimate long-term effects of the pandemic and its transformative dimensions are still unfolding, but as in writing at the end of 2021, there's a growing consensus that there will be no return to "normal." COVID-19 is proving transformative even though much of the response to the pandemic attempted to contain its systems-altering significance. A major evaluation challenge looking ahead will be to track, document, and extract lessons from just how transformative the COVID-19 pandemic turns out to be.

A team of internationally recognized experts, including Nobel prize winner Joseph Stiglitz and well-known climate economist Nicholas Stern, came together to assess the economic and climate impact of taking a green route out of the pandemic crisis. They catalogued more than 700 stimulus policies into 25 broad groups and conducted a global survey of 231 experts from 53 countries, including senior officials from finance ministries and central banks. Their analysis of whether COVID-19 fiscal recovery packages will accelerate or retard progress on climate change portrays the interconnection between the COVID-19 pandemic, economic policies, and environmental consequences which, taken together, illustrate the transformations necessary to attain a more sustainable and equitable future (Hepburn et al., 2020). Follow-up analysis asked: Are we building back better (O'Callaghan & Murdock, 2021)?

The global climate conference (COP26) held in Glasgow in November 2021 spotlighted the need for global, longer term sustainability transformation. The global health emergency is a short-term crisis within the larger and longer-term global climate emergency. This health crisis has revealed both the importance and possibility of systems transformation. This crisis illuminates the scale, scope, and urgency of systems transformations needed worldwide to create a more sustainable and equitable future. This pandemic is reflecting the fragmented and fragile nature of current systems, inadequate if we want a just and equitable world.

Evaluating transformation requires solid ethical grounding. As previously noted, the ethics of transformation involve interconnections between personal ethics (transforming our own behaviors), professional ethics (actively advocating a transformational stance among professional evaluators), societal

ethics (examining evaluation's role supporting the public good and democratic processes), and global ethics (ensuring attention to and engagement with the global emergency by incorporating transformational criteria of equity and sustainability into all evaluations based on human rights).

We each have a personal and professional responsibility to reflect on our role in transformation. As your work adapts to the current reality, think about how you can bring transformational perspectives to bear in it, to be attentive to it, gather evidence for, and support the kinds of transformations that may be needed after the pandemic subsides. Balancing long-term threats to the future of humanity with the urgent demands of short-term, crisis-generated interventions demands in-depth transformative evaluative thinking. Evaluators, individually and collectively, need to be prepared to contribute to finding and following pathways and trajectories toward transformations for a more equitable and sustainable future.

References

Bamberger, M., & Segone, M. (2011). *How to design and manage equity-focused evaluations*. UNICEF.

Bamberger, M., Vaessen, J., & Raimondo, E. (2016). *Dealing with complexity in development evaluation: A practical approach*. Sage.

Berg, R. V., Magro, C., & Adrien, M.-H. (Eds.). (2021). *Transformational evaluation for the global crises of our times*. International Development Evaluation Association. https://ideas-global.org/ideas-book-transformational-evaluation-for-the-global-crises-of-our-times/

Berg, R. V., Magro, C., & Mulder, S. S. (Eds.). (2019). *Evaluation for transformational change: Opportunities and challenges for the sustainable development goals*. International Development Evaluation Association. https://ideas-global.org/wp-content/uploads/2019/11/2019-11-05-Final_IDEAS_EvaluationForTransformationalChange.pdf

Bitar, K. (2021, February). *A social equity assessment tool (SEAT) for evaluation*. ResearchGate. https://www.researchgate.net/publication/349623379_A_social_equity_assessment_tool_SEAT_for_evaluation

Campbell-Patton, C. E., & Patton, M. Q. (2022). *Utilization-focused evaluation: Premises and principles* (5th ed.). Sage.

Chelsky, J., & Kelly, L. (2020, April 1). Bowling in the dark: Monitoring & evaluation during COVID-19 (coronavirus). *Independent Evaluation Group*. https://ieg.worldbankgroup.org/blog/mande-covid19

Chouinard, J. A. & Hopson, R. (2016) Canadian Journal of Program Evaluation 30(3), 248–276.

Dean-Coffey, J. (2018). What's race got to do with it? Equity and philanthropic evaluation practice. *American Journal of Evaluation*, *39*(4), 527–542. https://doi.org/10.1177%2F1098214018778533

Donaldson, S & Picciotto, R. (Eds.) (2016) Evaluation for an equitable society. IAP

Dubb, S. (2020, January 16). *Amnesty and planned parenthood identify their top policy priorities*. Nonprofit Quarterly. https://nonprofitquarterly.org/amnesty-and-planned-parenthood-identify-their-top-policy-priorities/

Duhigg, C. (2020, April 26). Seattle's leaders let scientists take the lead. New York's did not. *The New Yorker*, May 4, 16–22. https://www.newyorker.com/magazine/2020/05/04/seattles-leaders-let-scientists-take-the-lead-new-yorks-did-not

Equitable Evaluation Initiative. (2021). *Equitable evaluation framework*. https://www.equitableeval.org/ee-framework

Furubo, J.-E., Rist, R. C., & Speer, S. (Eds.). (2013). *Evaluation and turbulent times: Reflections on a discipline in disarray*. Routledge.

Global Alliance for the Future of Food. (2020). *Transforming food systems to improve human, animal, and planetary health*. https://futureoffood.org/our-approach/systems-thinking/

Heider, C. (2017). *Rethinking evaluation*. International Bank for Reconstruction and Development and The World Bank. http://ieg.worldbankgroup.org/sites/default/files/Data/ehinkingEvaluation.pdf

Hepburn, C., O'Callaghan, B., Stern, N., Stiglitz, J., & Zenghelis, D. (2020). Will COVID-19 fiscal recovery packages accelerate or retard progress on climate change? *Oxford Review of Economic Policy*, *36*(Supplement_1), S359–S381. https://doi.org/10.1093/oxrep/graa015

Hodgson, A. (2020). *Systems thinking for a turbulent world: A search for new perspectives*. Routledge.

Hood, S., Hopson, R., & Frierson, H. (Eds.). (2015). *Continuing the journey to reposition culture and cultural context in evaluation theory and practice*. Information Age Publishing.

Independent Evaluation Group. (2016). *Supporting transformational change for poverty reduction and shared prosperity – Lessons from the World Bank experience*. World Bank. https://ieg.worldbankgroup.org/sites/default/files/Data/Evaluation/files/WBGSupportTransformationalEngagements.pdf

Independent Evaluation Group. (2018). *Evaluation of GEF support for transformational change*. Global Environmental Facility. http://gefieo.org/sites/default/files/documents/reports/transformational-change.pdf

Independent Evaluation Group. (2020). *Evaluative resources and lessons to inform the COVID-19 response*. The World Bank. http://ieg.worldbankgroup.org/topic/covid-19-coronavirus-response?utm_source=Independent+Evaluation+Group+Contacts&utm_campaign=fee9c85338-EMAIL_CAMPAIGN_2020_05_04_05_44_COPY_03&utm_medium=email&utm_term=0_fc6e7f2a32-fee9c85338-113584913

International Development Evaluation Association. (2019). *Prague declaration on evaluation for transformational change*. https://ideas-global.org/wp-content/uploads//2019/10/Prague-Declaration-4-October-2019.pdf

Johnson, A. R. (2019, June 25). A look at language week: (On the absence of) whiteness. *AEA365*. https://aea365.org/blog/a-look-at-language-week-on-the-absence-of-whiteness-by-a-rafael-johnson/

Julnes, G. (Ed.). (2019). Special issue – Evaluating sustainability: Evaluative support for managing processes in the public interest. *New Directions for Evaluation*, *2019*(162). https://onlinelibrary.wiley.com/toc/1534875x/2019/2019/162

Lydenberg, S., & Burckart, B. (2020). *Assessing system-level investments: A guide for asset owners*. The Investment Integration Project. https://www.tiiproject.com/wp-content/uploads/2020/04/Assessing-System-Level-Investments_FINAL_04-21-2020.pdf

McKegg, K. (2019). White privilege and the decolonization work needed in evaluation to support Indigenous sovereignty and self-determination. *Canadian Journal of Program Evaluation*, *34*(2), 357–367. https://doi.org/10.3138/cjpe.67978

Mertens, D. M. (1999). Inclusive evaluation: Implications of transformative theory for evaluation. *American Journal of Evaluation*, 20(1), 1–14. https://doi.org/10.1177%2F109821409902000102

Mertens, D. M. (2009). *Transformative research and evaluation*. Guilford.

Mukherjee, S. (2020). What the coronavirus reveals about American medicine. *The New Yorker*, May 4, 24–31. https://www.newyorker.com/magazine/2020/05/04/what-the-coronavirus-crisis-reveals-about-american-medicine

O'Callaghan, B. J., & Murdock, M. (2021). *Are we building back better?* United Nations Environment Programme. https://wedocs.unep.org/bitstream/handle/20.500.11822/35281/AWBBB.pdf

Ofir, Z. (2017, October 16). The DAC evaluation criteria, part 5 – Non-negotiable criteria. *Evaluation for Development*. http://zendaofir.com/updating-dac-evaluation-criteria-part-5/

Ofir, Z. (2018a, July 31). Zenda's top YEE tips 2. *Evaluation for Development*. http://zendaofir.com/zendas-top-ten-tips-for-yees-2/

Ofir, Z. (2018b, August 20). Made in Africa evaluation 2 – Africa-rooted evaluation. *Evaluation for Development*. https://zendaofir.com/made-africa-evaluation-part-2-evaluation-rooted-africa/

Ofir, Z. (2018c, August 23). Made in Africa evaluation 4 – The NICE framework. *Evaluation for Development*. https://zendaofir.com/the-nice-framework-part-1/.

Olazabal, V. (2021, November 9). *Presidential plenary remarks*. American Evaluation Association conference, virtual.

Patton, M. Q. (2011). *Developmental evaluation: Applying systems thinking and complexity concepts to enhance use*. Guilford.

Patton, M. Q. (2020a). *Blue marble evaluation: Premises and principles*. Guilford Press.

Patton, M. Q. (2020b, March 23). Evaluation implications of the coronavirus global health pandemic emergency. *Blue Marble Evaluation*. https://bluemarbleeval.org/latest/evaluation-implications-coronavirus-global-health-pandemic-emergency

Patton, M. Q. (2021a). Evaluation criteria for evaluating transformation: Implications for the coronavirus pandemic and the global climate emergency. *American Journal of Evaluation*, *42*(1), 53–89. https://doi.org/10.1177%2F1098214020933689

Patton, M. Q. (2021b). How far dare evaluators go in changing the world? *American Journal of Evaluation*, *42*(2), 162–184. https://doi.org/10.1177%2F1098214020927095

Patton, M. Q. (2022). Evaluation ethics in desperate times. In R. D. van den Berg, P. Hawkins, & N. Stame (Eds.), *Ethics for evaluation* (pp. 235–255). Routledge.

Patton, M. Q., Podems, D., & Wildschut, L. (2021). *Synthesis of independent dialogues, Report 3*. UN Food Systems Summit. United Nations. https://www.un.org/sites/un2.un.org/files/unfss_independent_dialogue_synthesis_report_3_0.pdf

Rasmussen, S. A., & Goodman, R. A. (Eds.). (2018). *The CDC field epidemiology manual*. Centers for Disease Control and Prevention. https://www.cdc.gov/eis/field-epi-manual/index.html

Richardson, R.A. (2021) Global Alliance Principles. In Richardson, R. A., & Patton, M. Q. (2021). Leadership-evaluation partnership: Infusing systems principles and complexity concepts for a transformational alliance. New Directions for Evaluation, No. 170, 139-147.

Rowe, A. (2019). Evaluating sustainability: Evaluative support for managing processes in the public interest. *New Directions for Evaluation*, 2019(162), 1–154. https://onlinelibrary.wiley.com/toc/1534875x/2019/2019/162

Rugg, D., Buehler, J., Renaud, M., Gilliam, A., Heitgerd, J., Westover, B., Wright-Deageuro, L., Bartholow, K., & Swanson, S. (1999). Evaluating HIV prevention: A

framework for national, state and local levels. *American Journal of Evaluation*, *20*(1), 35–56. https://doi.org/10.1177/109821409902000104

Segone, M., & Tateossian, F. (2017). No one left behind. In M. Bamberegr, M. Segone, & F. Tateossian (Eds.). *Evaluation for agenda 2030: Providing evidence on progress and sustainability* (pp. 23–34). https://ideas-global.org/wp-content/uploads/2017/12/Chapter-2.pdf

Shanker, S. (2019). Definitional tension: the construction of race in and through evaluation. [Doctoral dissertation, University of Minnesota]. University Digital Conservancy. https://hdl.handle.net/11299/211799

Shanker, V. (2019, June 24). A look at language week: Tidying evaluation of "diversity" & "culture." *AEA365*. https://aea365.org/blog/a-look-at-language-week-tidying-evaluation-of-diversity-culture-by-vidhya-shanker/

Smith, L. T. (1999). *Decolonizing methodologies: Research and indigenous peoples*. Zed.

Stake, B. (2004). How far dare an evaluator go toward saving the world? *American Journal of Evaluation*, *25*(1), 103–107. https://journals.sagepub.com/doi/pdf/10.1177/109821400402500108

Sturdevant, L. (2021, November 20). Minnesota, the midwest are central in combating climate change. *Star Tribune*. https://www.startribune.com/minnesota-the-midwest-are-central-in-combating-climate-change/600119044/

Systems in Evaluation. (2018). *Principles for effective use of systems thinking in evaluation*. Systems in Evaluation Topical Interest Group, American Evaluation Association. https://www.betterevaluation.org/en/resources/principles-effective-use-systems-thinking-evaluation#:~:text=%E2%80%9CTo%20apply%20the%20%E2%80%9Coverarching%E2%80%9D,%2C%20perspectives%2C%20boundaries%20and%20dynamics

The Investment Integration Project. (2020, May 12). *Who is doing well, who isn't: Evaluating the performance of asset managers against systemic social and environmental progress* [Online panel discussion].

Tufekci, Z. (2021, November 21). After a pandemic failure, the US needs a new public spirit. *New York Times Sunday Review*. https://www.nytimes.com/2021/11/18/opinion/covid-winter-risk.html

Uitto, J. I. (2019). Sustainable development evaluation: Understanding the nexus of natural and human systems. *New Directions for Evaluation*, 162, 49–67. https://doi.org/10.1002/ev.20364

United Nations (2015). Transforming our world: the 2030 agenda for sustainable development. Transforming our world: the 2030 Agenda for Sustainable Development | Department of Economic and Social Affairs (un.org)

United Nations. (2021, November 1). *COP26: Enough of 'treating nature like a toilet' – Guterres brings stark call for climate action in Glasgow*. UN News. https://news.un.org/en/story/2021/11/1104542

United Nations Children's Fund. (2015). *Unless we act now: The impact of climate change on children*. https://www.unicef.org/reports/unless-we-act-now-impact-climate-change-children

United Nations Evaluation Group. (2016). *Norms and standards for evaluation*. United Nations Evaluation Group. http://www.unevaluation.org/document/download/2787

Weiss, C. H., & Connell, J. P. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. In P. Connell, A. C. Kubisch, L. B. Schorr, & H. Weiss (Eds.), *New approaches to evaluating community initiatives: Concepts, methods, and contexts* (pp. 65–92). The Aspen Institute.

Afterword

Ray C. Rist

As I write this Afterword in the summer of 2022, the global death toll from the COVID-19 virus is now estimated at more than 15 million. And this estimate is only partial.[1] The actual number is believed to be much higher, possibly two times higher. A recent edition of the medical journal *The Lancet* (Wang et al., 2022) concluded that an estimated 18.2 million persons have perished from this virus, excluding data from India, which is refusing to release information on its own deaths from COVID-19. In many parts of the world, people do not have access to hospitals, medical care, testing, or health professionals for treatment. Moreover, the vaccines are not equally accessible across, let alone within, all countries. It is an uneven picture, superimposed on what is now the biggest health and humanitarian crisis in more than a generation.

The nine papers included in this book cover issues addressed from four countries on three continents, and from the United Nations as an organization, and were all written toward the end of the second year of the pandemic. What follows is a brief set of five observations that deserve evaluators' attention as we now move into the third year of the pandemic. These observations are brief as the crisis is still underway, and one needs to be careful about making definitive statements in what is clearly a continuously evolving situation. We can observe trends and tendencies, but the world is still in the midst of illness, death, and new infections every day. We must be cautious and circumspect because it is the right thing to do. Broad pronouncements are neither warranted nor required, but ongoing careful evaluative study is mandated. Thus, the brevity.

Furthermore, as Joseph Stiglitz (2020) has written, "Even as we emerge from this crisis, we should be aware that some other crisis surely lurks around the corner. We can't predict what the next one will look like – other than it will look different from the last."

Observation one

An emergent trend from the pandemic, now into its third year, is that it is strengthening the institutionalization and reach of "big" government. This stands in contrast to several previous decades in which the role of

government has been reduced in size and scope as societies have increasingly relied on the market. As Schwab and Malleret (2020, p. 31) have observed: "This is a situation that is set to change because it is hard to imagine how an exogenous shock of such magnitude as the one inflicted by COVID-19 could be addressed with purely market-based solutions. Already and almost overnight, the coronavirus succeeded in altering perceptions about the complex and delicate balance between the private and public realms in favor of the latter."

So how will we know that government is indeed expanding and pushing the markets into a lesser role? The key means of knowing is to examine the relationship of government to the economy. The intervention of governments has not been painless, but it has been quick. It has also been without precedent since Second World War. The scale of stimulus programs in the trillions of dollars to support the welfare of citizens and maintain employment cannot be overemphasized. This support has not come from the markets. It has come from governments. Only governments have the resources to do what was needed and without government intervention, we would be in a much deeper mess. This observation is worthy of careful investigation and deeper analysis.

Observation two

The pandemic has exposed some deep structural flaws and frailties in our societies. These include weaknesses in how we treat the elderly, the poor, minorities, children, and our response to the care economy. All of these are contexts and populations in which we have not created an appropriate level of societal responsibility, nor responded adequately to their individual needs. Indeed, it can be argued that persons in these groups (those persons most often in the lower income brackets) were on the first line of defense during the pandemic, but were overlooked, ignored, and left more vulnerable to COVID-19 infections that rampaged through our countries. They were also most frequently omitted from early access to the vaccines, leading to disproportionately high death rates. As a result, we have broken the tenuous social contract with persons in these groups. This is most clear in relation to how we address and respond to the broad issues of inequality and marginalization. How we choose to respond as individual societies will be one of the defining characteristics in the post-pandemic era. This topic also merits close attention from evaluators in the months and years to come.

Observation three

The COVID-19 crisis has exposed the glaringly inadequate status of many, if not most national health systems. For the United States, for example, to have thus far lost more than one million lives is simply so unacceptable as

to be unspeakable. For one of the wealthiest nations on earth to have tolerated this level of death and disease is so far from being defensible as to be immoral. This links in a profound way back to the proposed Observation two – a recognition of deep structural flaws in how we presently organize our societies. Again, how we propose to respond will be among the most defining characteristics of the post-pandemic era. At present, it is evident that the pandemic is no great leveler. Evaluators can dig deep into seeking to understand this set of inequalities.

Observation four

The cost of misinformation is considerable. "Freaking miracle" – this is how health journalist Helen Branswell (2022) describes the vaccines that that saved millions of lives from COVID-19. The vaccines, developed and offered in countless countries, have proven to be 90 percent effective against the infection. They are safe and – also relevant – they are often free.

But sadly there are many who think otherwise. They hold to conspiracy theories about the vaccines. A recent survey of 18,782 persons across all 50 states of the United States posed four vaccine misinformation claims, asking respondents if the claim was true or false or if the person was not sure. True or false: The vaccines contain microchips? Five percent said yes. The vaccines contain aborted fetal cells? Seven percent checked true. The vaccines can alter human DNA? Eight percent agreed. The vaccines can cause infertility? Ten percent thought this to be true. A full 46 percent were uncertain about the veracity of one or more of the four statements.

Misinformation about vaccines is directly correlated with whether persons get the jab. The survey showed that among those who did not believe any of the false statements, 80 percent were already vaccinated. In the group that thought one or more false statements were true, 60 percent have not been vaccinated. And the data show persons who do not get at least the first shot are 14 times more likely to die from COVID-19 than the vaccinated.[2] Thus the question: How many of the more than 1,006,000 deaths in the United States could have been averted with the vaccine? Clearly, the answer is: "many!" Studies of the resistance to the shots by demographics, religion, politics, and socioeconomic status are all fruitful areas for evaluative work to start to unpack the causal connections.

Observation five

Many commentators have spoken and written of eventually going back to normal, however "normal" is defined. But this is not possible. The turmoil of COVID-19 will not be reversed in its entirety. For example, the tension between the economy and public health will not disappear, particularly as some will argue it is acceptable to sacrifice a few (in this case the elderly)

for the sake of the economy, a shocking position that has been expressed by Texas Lieutenant Governor Dan Patrick (Knodel, 2020). This is a false (and potentially deadly) trade-off.

So, dear colleagues, I recommend this book to you while we are all still in the process of working through the implications of this vast challenge to our lives and our societies. The virus is not yet through with us. And even when that time comes, as Stiglitz (2020) notes, another crisis will grab our attention, our pocketbooks, and even our lives. We need to stay alert and prepared. Because COVID-19 is an example of what happens when we are neither alert nor prepared.

Notes

1 This point is discussed more extensively in an editorial entitled "Millions More Lost," Editorial, *The Washington Post*, March 14, 2022 (p. A18), and again on April 17, 2022.
2 This point is discussed more extensively in an editorial entitled, "The Cost of Misinformation," *The Washington Post*, February 23, 2022.

References

Branswell, H. (2022, February 14). *Why covid-19 vaccines are a freaking miracle*. STAT. https://www.statnews.com/2022/02/14/why-covid-19-vaccines-are-a-freaking-miracle/

Knodel, J. (2020, March 24). *Texas Lt. Gov. Dan Patrick suggests he, other seniors willing to die to get economy going again*. NBC News. https://www.nbcnews.com/news/us-news/texas-lt-gov-dan-patrick-suggests-he-other-seniors-willing-n1167341

Schwab, K., & Malleret, T. (2020). *COVID-19: The great reset*. Forum Publishing.

Stiglitz, J. (2020, April 8). A lasting remedy for the covid-19 pandemic's economic crisis. *The New York Review*. https://www.nybooks.com/daily/2020/04/08/a-lasting-remedy-for-the-covid-19-pandemics-economic-crisis/

Wang, H., Paulson, K. R., Pease, S. A., Watson, S., Comfort, H., Zheng, P., Aravkin, A. Y., Bisignano, C., Barber, R. M., Alam, T., Fuller, J. E., May, E. A., Jones, D. P., Frisch, M. E., Abbafati, C., Adolph, C., Allorant, A., Amlag, J. O., Bang-Jensen, B., … Murray, C. J. L. (2022). Estimating excess mortality due to COVID-19 pandemic: A systematic analysis of COVID-19-related mortality, 2020–21. *The Lancet*, *399*(10334), 1513–1536. https://doi-org.proxy3.library.mcgill.ca/10.1016/S0140-6736(21)02796-3

Index

Note: *Italicized* and **bold** pages refer to figures and tables respectively, and page numbers followed by "n" refer to notes.