

Evaluation as a Tool for Research, Learning and Making Things Better

Evaluation as a Tool for Research,
Learning and Making Things Better

Edited by

Satu Kalliola

CAMBRIDGE
SCHOLARS

P U B L I S H I N G

Evaluation as a Tool for Research, Learning and Making Things Better,
Edited by Satu Kalliola

This book first published 2014

Cambridge Scholars Publishing

12 Back Chapman Street, Newcastle upon Tyne, NE6 2XX, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Copyright © 2014 by Satu Kalliola and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-4438-5637-1, ISBN (13): 978-1-4438-5637-9

CONTENTS

List of Illustrations	ix
List of Tables	xi
Acknowledgements	xiii

Introduction Satu Kalliola Evaluation: Methodology for Humanitarian Innovators	1
--	---

PART I: DIFFERENT APPROACHES TO EVALUATION METHODS AND DESIGNS

CHAPTER ONE

Edward J Mullen: Comparative Effectiveness Research: Designs and Methods for Social Intervention Evaluation	13
---	----

CHAPTER TWO

Mansoor Kazi: Realist Evaluation and Effectiveness Research: An Example From School-Based Interventions	31
---	----

CHAPTER THREE

Brendan McCormack and Catherine Buckley: Holistic and Authentic Evaluation in Health Care	45
--	----

CHAPTER FOUR

Ross VeLure Roholt and Michael Baizerman: “Making Things Better” by Using Evaluation Advisory Groups ...	69
---	----

CHAPTER FIVE

Rogério Gimenes Giugliano:

Índice de Gestão Descentralizada and the *Bolsa Familia* Programme:

Evaluating and Developing Brazilian Municipalities' Management

Capabilities 87

PART II: THE EVIDENCE-BASED PRACTICES IN HEALTH AND SOCIAL CARE

CHAPTER SIX

Eamonn Noonan:

Perspectives on Evidence-Based Practice in Social Care 109

CHAPTER SEVEN

Maritta Välimäki:

Evidence-based Practice as a Tool for Researching,

Learning and Making Improvements in Healthcare 127

CHAPTER EIGHT

Gavin Heron:

Social Work Education: A Framework for Evaluating

Feedback on Student Social Workers' Understanding

of Child Care and Protection 143

PART III: EVALUATION OF THE DEVELOPMENT OF WORKING LIFE

CHAPTER NINE

Robert Arnkil:

Bridges Over Troubled Evaluation Waters 161

CHAPTER TEN

Ijin Hong and Madhu Sudhan Atteraya:

Evaluation and Migration:

Qualitative Perspectives on Migrant Workers in South Korea 179

CHAPTER ELEVEN

Kevin D. Moore and Assumpta Ryan:
“To Keep a Person in their Own Wee Corner”:
Evaluating the Role of Home Care Workers in Health
and Social Care Using a Grounded Theory Approach 197

PART IV: EVALUATIONS IN THE DIFFERENT LEVELS OF EDUCATION

CHAPTER TWELVE

Giovanni Abbiati, Gianluca Argentin and Aline Pennisi:
Learning from Implementation: The Case of the
Evaluation of a Professional Development Programme
for Mathematics Teachers in Italy 223

CHAPTER THIRTEEN

Anu Räisänen and Mari Rökköläinen:
Developmental Assessment of Learning Outcomes 241

CHAPTER FOURTEEN

Shelley Kinash, James Herbert and Romy Lawson:
Why so Evasive about Evaluation in Higher Education? 267

List of Contributors 287

LIST OF ILLUSTRATIONS

Figure 1-1. CER Types and Methods	21
Figure 1-2. PCORI Translation Framework Component.....	25
Figure 2-1. Jamestown Senior High Average Cumulative School Grade School 2008–09.....	36
Figure 3-1. Creating a landscape of my workplace culture.	58
Figure 3-2. Creative Hermeneutic Data Analysis	60
Figure 5-1. Evolution of the percentage of families under the higher poverty line of the Millennium Goals in Brazil since 1995.	92
Figure 5-2. Percentage of Brazilian population below the extreme poverty line.....	92
Figure 5-3. The formula for calculating the IGD.....	94
Figure 5-4. The graph displays the mean IGD over the duration of the study.....	98
Figure 5-5. IGD of all Brazilian Municipalities at the beginning of the Research in April 2006.....	99
Figure 5-6. IGD of all Brazilian Municipalities at the end of the research in November 2007	99
Figure 5-7. Evolution of Brazilian Municipalities according to population size.	100
Figure 5-8. Evolution of municipalities according to Human Development Index.	101
Figure 8-1. Frequency of the seven principles in relation to child care and protection issues.....	152
Figure 9-1. Disconnects of learning, evidence and evaluation.	164
Figure 9-2. Microcosms learning from each other in a peer-learning workshop.....	171
Figure 9-3. The proposed “bridge”: A meso-level dialogical learning space.....	175
Figure 10-1. Evaluation designs.....	184
Figure 11-1. The Conceptual Paradigm Model: The Home Care Worker Role, Illustrating Relationship of Major Categories to Each Other and to the Core Category	203
Figure 13-1. Assessment of learning outcomes for the vocational qualification in electrical and automation engineering.	247
Figure 13-2. Targets of competence assessment and the transfer of competencies.....	254

Figure 13-3. Factors related to the dependencies of learning
outcomes. 255

LIST OF TABLES

Table 2-1. The binary logistic regression model for those achieving schoolgrades of 70% or less and those achieving greater than 70%.....	37
Table 3-1. A framework for authentic evaluation.....	49
Table 5-1. The number of families involved in PBF, and Federal Investments in the programme.	91
Table 5-2. The responsibilities of the three government levels in the <i>Bolsa Familia</i> Programme.	93
Table 9-1. Basic strategies and methods of dissemination.	167
Table 9-2. Human well-being and evaluation.	168
Table 9-3. Key characteristics of the <i>peer-learning workshop</i>	173
Table 10-1. Advantages of qualitative research approaches for evaluation: a case study.	192
Table 12-1. The behaviour of non-compliers.	231
Table 12-2. Reasons for leaving/not showing up for the course	232
Table 13-1. System of student assessment and assessment of learning outcomes.	245
Table 13-2. Strengths and problems involved in the assessment of learning outcomes, based on vocational skills demonstrations in terms of their reliability	249

ACKNOWLEDGEMENTS

Editor Satu Kalliola

This book is a result of co-operation between an international group of enthusiastic and positively ambitious evaluation researchers and practitioners. The foundation for the book was laid in the summer 2012, in the closing ceremony of The 8th International Conference on Evaluation for Practice Eval2012, held in the City of Pori, Finland. An idea came up to compile all the Plenary Presentations to a book, complemented by a few of the most interesting workshop presentations. All the presenters in question accepted the invitation to develop their conference papers to full articles.

I gladly use this opportunity to thank the authors of this book who saw the effort worthwhile.

Also, I find it important to thank once more all the sponsors of Eval2012. The support of Tampere University School of Social Science and Humanities, Tampere University Foundation, University Consortium of Pori, Federation of Finnish Learned Societies, and The Finnish Work Environment Fund.

My final thanks go to our Commissioning Editor, Carol Koulikourdi, who first appreciated the potential of this volume and who has helped us through the publication process.

Authors of Chapter Twelve:

Learning from Implementation: The Case of the Evaluation of a Professional Development Program for Math Teachers in Italy

Giovanni Abbiati, Gianluca Argentin, and Aline Pennisi

We deeply benefited from the work of Andrea Caputo (INVALSI) and Daniele Vidoni (INVALSI). Special thanks also to: Alberto Martini (who contributed to design and set up the project), Piero Cipollone, Annamaria Leuzzi (Ministry of Education), Annamaria Fichera (Ministry of Education), Alessandro Cavalli, Maria Pia Perelli d'Argenzio, Enrico Rettore, Jaap Scheerens. M@t.abel<mailto:M@t.abel> is an acronym for *Matematica, apprendimenti di base con e-learning*. We evaluate here the most recent version of this program, carried out with EU funding - PON Istruzione 2007-2013 (I-3-FSE-2009-2).

Author of Chapter Seven:
Evidence-based Practice as a Tool for Researching, Learning,
and Making Things Better

Maritta Välimäki

Writing this book chapter was funded by Finnish Cultural Foundation, Foundations' Professor Pool, and the University of Turku, which are greatly appreciated.

INTRODUCTION

EVALUATION:
METHODOLOGY FOR HUMANITARIAN INNOVATORS

SATU KALLIOLA

Evaluation as a Scientific Discipline

Scientific disciplines have certain intertwined characteristics in common: they have academic societies, conferences and journals, all of which exist at both national and international levels. Usually, such academic societies organize conferences and publish journals to introduce, test and disseminate new research results. Evaluation meets all of these criteria, but it tends to be seen as a special *methodology* of social research rather than a discipline. A distinct feature of evaluation as a scientific discipline is the involvement of practitioners who both require and use the results of evaluation.

It can be argued that evaluation has created different schools, similar to that of practice-oriented research traditions and particular social research methodologies. These schools debate the most efficient ways of collecting and analysing data on the benefits of human services and other social interventions, such as new social and labour legislation, or education. Methodological arguments are presented on how to best serve people in need. The pursuits of service providers, for example governmental or voluntary organizations and other agencies, are also taken into account.

From the viewpoint of rigorous research, the issue of how to best serve people in need concerns our assumptions and hypotheses about human existence and the nature of knowledge. In the history of evaluation, experimental models and quantitative research methods with objectivity as their final aim have been prominent. This type of classic social research has gained even more favour since the idea of collecting evidence for best practices diffused from medicine and health services to other human services. It can be argued that today evidence-based practice development in most human services owes a debt to a voluntary research network called the Cochrane Collabo-

significant book *Effectiveness and Efficiency – Random Reflections on Health Services* was first published. Since then, the Cochrane Collaboration has grown, receiving both plaudits and criticism, but has continued to be devoted to better health care, and it guides researchers, care providers and patients in the right direction by conducting systematic reviews of primary research (www.cochrane.org/about-us/history). Nowadays, the Cochrane Collaboration co-operates with the Campbell Collaboration, an international research network that produces systematic reviews in the fields of education, crime and justice, social welfare and international development (www.campbellcollaboration.org/about-us).

In addition to systematic reviews, there are other strong methodological approaches that contribute to evaluation as a scientific discipline. Realistic evaluation by Pawson and Tilley (1997) and qualitative approaches by Patton (2002) have grown to become institutions in their own right. During the course of realistic – later realist – evaluation, researchers, service providers, practitioners and beneficiaries engage in an exciting search for theories about the mechanisms and contexts that enable desired change. Following Patton (2002), evaluation researchers have broadened the scope of their methodological choices and have learned to decide when it is appropriate to use different and complementary data collection methods. Yet another significant approach to evaluation comes from Guba and Lincoln (1989), who present evaluation as a constructivist enquiry with a participatory orientation.

Lately, the classic evaluation strategies have been challenged by new evolving types of intervention evaluation, for example comparative effectiveness research (Velengtas, Mohr, and Messner 2012) that has started to be acknowledged by political decision-makers in the development of new legislation in the USA. In addition, many evaluation researchers select their methods depending on the problem to be solved and combine research methods. On the other hand, it is probable that the vast majority of human services are delivered without any evaluation, and new laws regulating service delivery are passed without prior reference to the anticipated, research-based outcome.

Although a large part of the results of evaluation research is reported only to the stakeholders or commissioners of evaluations, there is a vast and probably growing number of scientific journals that concentrate on evaluation. For example, Sage, a publishing house specializing in social and related sciences, publishes seven journals with “evaluation” in the title.

Numerous national and international conferences on evaluation are arranged every year. Suomen arviointiyhdistys – Finnish Evaluation Society,

as a member of the European Evaluation Society, arranges conferences on a regular basis. In addition, there are a number of networks specializing in the evaluation of specific themes (www.sayfes.fi; www.europeanevaluation.org).

This book is a product of an International Conference on Evaluation for Practice – a conference series that takes seriously the involvement of practitioners and the improvement of human services through the use of evaluation results. The active organizers are mostly university scholars who have permanent contacts with evaluation users. Since the first conference in 1995 in the UK, the conference has been organized in Finland twice, in 2002 and again 10 years later. The 8th International Conference on Evaluation for Practice was organized at the University Consortium of Pori in June 2012. The theme of the conference was “Evaluation as a Tool for Research, Learning, and Making Things Better”.

Wanting to Make Things Better

The organizers and participants of International Conferences on Evaluation for Practice are very aware of the ever present organizational and political contexts of evaluation. The issues are the same as in other fields of applied social sciences; social life offers a multitude of interesting research topics, but there are always constraints hindering the direct use of research results.

As the chairperson of the 8th conference and the editor of this book, which is compiled from the work of the Plenary Speakers, Plenary Panel and outstanding workshop presentations, I would like to argue that there exists a large number of evaluation researchers who are passionate for their research to have a positive impact on the lives of the people they have studied. They are able to fruitfully combine their scientific ambitions of making things better with learning from evaluations and amending programmes and other social interventions according to evaluation results. The spirit is the same as that of Kurt Lewin, who coined the term “action research”. Lewin (1948) argued that research that produces nothing but scholarly books will not suffice. He described action research as a spiral staircase; each step creates a circle of planning, action and fact-finding about the results of action (Lewin 1948), with a central focus being that of evaluation and learning from evaluation. It is also important to note that in action research, planning is preceded by fact-finding about the initial situation, the problem to be solved, or harmful conditions to be ameliorated.

The basic elements of many evaluation designs are the same as in the Lewinian action research tradition which also has clear connections to the

notions of Rossi and Freeman (1991). For Rossi and Freeman (1991, 13–14) evaluation research is a robust arena of activity directed at collecting, analysing, and interpreting information on the necessity, implementation, effectiveness and efficiency of intervention efforts to the lot of humankind. They continue by describing more concretely the object of evaluation research at the different stages of making social interventions: in the planning stages, evaluation focuses on the extent and severity of the problems requiring social interventions, and on the design of programmes to ameliorate them; in the implementation stage, evaluation helps to determine the degree to which the interventions succeed in reaching their intended target populations and provide the resources, services, and benefits envisioned by the service providers; impact assessments estimate the effects and magnitudes of interventions; and, for decisions to continue, expand or curtail the interventions, evaluations calculate costs in relation to benefits. Sometimes evaluations try to cover all of these stages, but often they concentrate only on some of them.

Doing evaluations and using the results is seldom as simple and smooth as presented above, however. The complexity of human experience, the diversity of values, political power relations and the whole organizational context of evaluations place specific demands on the skills of researchers. They have to master social interaction, negotiation, guidance, training or teaching, and also managerial skills in order to survive and to undertake rigorous research in environments characterized by value conflicts. These challenging aspects of evaluation are also included in the comprehensive idea of evaluation by Rossi and Freeman (1991, 14) who see that in all cases the aim of evaluation is to provide the most valid and reliable findings possible within political and ethical constraints and the limitations imposed by time, money, and human resources.

In advancing knowledge in the development of evaluation research methods, the articles in this book also to some degree deal with the constraints mentioned earlier. The scope of human services evaluated expands from social and health services to regional development, working life and education.

The Contents of the Book

The focus of the articles in *Part I – Different Approaches to Evaluation Methods and Designs* – is on the versatile methodology of evaluation.

Edward Mullen describes and analyses Comparative Effectiveness Research (CER) that has gained favour in the United States in connection

with new legislation in 2009–2010. The main features of CER are the use of comparison research designs, real world settings and client-centredness. CER has emerged from earlier research emphases on technology assessment, evaluation research, outcomes research, and evidence-based policy and practice. CER is presented to have a non-hierarchical view of evidence, judged against the criteria of relevance and responsiveness to its requirements.

In his chapter, Mansoor A. F. Kazi asks the classic question of realist evaluation (Pawson and Tilley 1997): “what works, for whom and in what contexts?” in connection with evidence-based practice. He searches for answers within the framework of mixed methods by introducing a new evaluation strategy that combines the traditions of epidemiology and effectiveness research in human services. He shows how data collected for purposes other than evaluation may also be used for this purpose, and how this type of work enhances the opportunities for collectively valuing evaluation itself. The key is evaluation applied as a part of daily practice, and in partnership between evaluators and human service agencies, in this case US school districts.

Brendan McCormack, a Plenary Speaker of the 8th International Conference on Evaluation for Practice, and his co-author Catherine Buckley, discuss the dominance of “evidence-based practice” in health care practice and look for alternative and complementary approaches to the evaluation of effectiveness, aiming at evidence-informed and person-centred practices. Building on the notion of the complexity of human experience, they adopted integrated and authentic approaches to the design of methodologies in their case-study that helped nurse leaders working in nursing homes to develop evidence-informed and person-centred cultures of care.

Ross VeLure Roholt and Michael Baizerman discuss the diverse roles of advisory groups in evaluations. They expand the significance of advisory groups from soliciting expertise and giving political legitimacy to enhancing programme improvement, accountability, and decision-making. Their evaluation cases are examples mainly from non-governmental organizations, including advisory groups in social and health services, ethnic groups (indigenous peoples) and museums.

Rogério Gimenes Giugliano gives an example of how rigorous evaluation methods can be combined as a tool that works on three levels of programme management – federal, local/municipal and citizens/beneficiaries. He presents the main findings of an exploratory study carried out with data on the first 18 months of implementation of the Decentralized Management Index (*Índice de Gestão Descentralizada*), a synthetic indicator that works

as an evaluation tool and, at the same time, determines the funds that the Brazilian Federal Government transfers to municipalities for the management of the *Bolsa Família* Programme.

Perhaps the most familiar approaches to combining evaluation with everyday life are *the evidence-based practices in health and social care*, and *Part II* presents these activities. Eamonn Noonan and Maritta Välimäki, both Plenary Speakers at the 8th International Conference for Evaluation for Practice, write about the importance of doing systematic reviews on earlier primary research to gather solid evidence for best service practices. Both articles also focus on the problems related to evidence based practices: Noonan presents the pursuits of the Campbell Collaboration and social care; and Välimäki presents the pursuits of the Cochrane Collaboration and healthcare.

Eamonn Noonan's article highlights a case that discusses the critical responses in Scandinavia to a Campbell systematic review of school-based programmes to prevent bullying, and then considers the place of this review in the broader political debate. As a conclusion, the article suggests an enhanced infrastructure, both to support rigorous impact evaluation and systematic review, and to gather and communicate evidence to practitioners.

According to Maritta Välimäki, the various definitions of evidence-based practice have three elements in common, research evidence based on high quality studies, staff experiences and patients' needs. She points out how difficult it is to read, comprehend, and apply the reviewed research results in the hectic daily work of nursing. Although different models have already been developed to facilitate the implementation of research findings in practice, a more systematic approach is still needed in research, education, and practice development to make things better.

Differing from the earlier articles, Gavin Heron's article examines the use of evaluation in primary research to generate an evidence base. His subject is social work education, which he argues lacks effective evaluation. His detailed focus is on creating useful evaluation tools to secure the standards of social work education in the UK, but his message is valid in other countries too. In his chapter, the special emphasis is on assuring social workers' understanding about child care and protection.

Evaluation of the development of working life, a new theme at the International Conferences on Evaluation for Practice, is the focus of the articles in *Part III*. Robert Arnkil's article "Bridges over troubled evaluation waters" is based on his plenary presentation at the conference. The contents of his article are at the same time deeply academic and personal reflections of the status quo of evaluation research in Finland and transnationally. The

substance of working life can be found in his examples of the multi-actor peer-learning workshop methodology and the issues of power, co-operation, trust and respect. His underlying theme is learning from evaluation.

With the Republic of Korea as their backdrop, Ijin Hong and Madhu Sudhan Atteraya offer a different view of working life in the global economy while researching how to combine the themes of evaluation with issues of new migration – a topic that is usually approached from the standpoint of employment and integration policies. At the same time, migrant workers can be seen as a vulnerable group of people with various needs. The authors suggest that qualitative analyses can be usefully employed for evaluating integration policies of migrant people in countries of new immigration.

Worldwide demographic trends, indicating an expansion in the elderly population and other societal changes, having a direct impact on the nature and type of social and health services, is the background for Kevin Moore and Assumpta Ryan, who write about the working life of home care workers in the UK context. The focus is on the self-perceived role of home care workers who see many challenges in the recognition of the importance of their role within the wider health and social care community.

Part IV of the book is dedicated to *evaluations in the different levels of education*: in primary school teaching, vocational education and training, and in higher education. Giovanni Abbiati, Gianluca Argentin and Aline Pennisi present an interesting case from southern Italy where a thorough analysis of the implementation of a seemingly ineffective professional development programme for mathematics teachers reveals constraints related to organizational or bureaucratic barriers, or the national mathematics curriculum. The authors emphasize the significance of learning from evaluation in re-designing the programme.

The article by Anu Räisänen and Mari Rökköläinen discusses the assessment of learning outcomes in vocational education and training (VET), and the ensuing methodological challenges in Finland, where evaluation of education has been systematically developed since the 1990s. The authors describe the legislation and the role of National Board of Education. The focus of the article is on the integration of formative and summative assessments, and the tensions between vocational skills demonstrations and the evaluation of competence development in the context of work-oriented VET.

In the concluding chapter of the section, and the whole book, the role of evaluation in higher education is described and analysed by Shelley Kinash, James Herbert and Romy Lawson in an article that highlights evaluations as delicate issues due to their often political aspects. The focus of this Austral-

ian contribution is on the definitions and ways to evaluate quality, graduate attributes, assurance of learning and student feedback in higher education.

Another Angle?

Since all the articles are very rich in their scope and content, the organization of this book could have been different, realized in many cases only by minor shifts in the focus of the writing process. For example, Ross VeLure Roholt and Michael Baizerman's advisory groups among indigenous people and Ijin Hong and Madhu Sudha Atteraya's needs assessment among migrant workers would have composed an interesting section focusing on anthropological approaches in evaluation. Another new angle could have been formed from the contributions by Giovanni Abbiati, Gianluca Argentin and Aline Pennisi, as well as that of Anu Räisänen and Mari Rökköläinen, which could have been grouped together with Rogerio Gimenes Giugliano's work to form a section describing the evaluation methods in use by national governments.

Since Robert Arnkil's article crosses the borders of working life evaluations, it contributes broadly to the methodology of evaluation. The same can be said about the contribution by Kevin D. Moore and Assumpta Ryan, who look for new ways to understand home care work. The search for proper methodologies can also be found in the article by Shelley Kinash, James Herbert and Romy Lawson, who consider how to evaluate quality in higher education. Gavin Heron's contribution on generating an evidence base for social work education is as much about education as it is about evidence, and Mansoor A. F. Kazi's mixed methods article is also about interventions aimed at improving the basic prerequisites for school attendance.

Other articles also seen as methodological could have been organized differently, for example by emphasizing the very explicit patient-, client-, or more generally human-centredness of the articles by Edward J. Mullen, and Brendan McCormack and Catherine Buckley. Systematic reviews of primary research, advocated by Eamonn Noonan and Maritta Välimäki, aim for better social and health care from the point of view of the clients, that is, ordinary citizens.

This alternative overview serves as an alternative table of contents, which may inspire the readers of this book to appreciate evaluation research for its diversity. At its best, evaluation is ambitious and rewarding research conducted in the service of the public good.

References

- Campbell Collaboration. 2013 “About Us.” Accessed 8 December, 2013. <http://www.campbellcollaboration.org/about-us/>.
- Cochrane, Archibald J. 2004. *Effectiveness and Efficiency. Random Reflections on Health Services. New Introduction by Chris Silagy*. London: The Royal Society of Medicine. (Originally published by the Nuffield Provincial Hospitals in 1972.)
- Cochrane Collaboration. 2013. “History.” Accessed 8 December, 2013. <http://www.cochrane.org/about-us/history>.
- European Evaluation Society. 2013. “European Evaluation Society Homepage.” Accessed 11 December, 2013. <http://www.europeanevaluation.org/>.
- Guba, Egon G. and Yvonna S. Lincoln. 1989. *Fourth Generation Evaluation*. Newbury Park: Sage.
- Lewin, Kurt. 1948. “Action Research and Minority Problems.” In *Selected Papers on Group Dynamics*, edited by Gertrude Weiss Lewin, 201–216. New York: Harper & Brothers Publishers.
- Patton, Michael Quinn. 2002. *Qualitative Research & Evaluation Methods*. 3rd Edition. Thousand Oaks, London, New Delhi: Sage.
- Pawson, Ray and Nick Tilley. 1997. *Realistic Evaluation*. London, Thousand Oaks, New Delhi: Sage.
- Rossi, Peter H., and Howard E. Freeman. 1991. *Evaluation. A Systematic Approach*. Fourth Edition. Newbury Park, London, New Delhi: Sage.
- Suomen arviointiyhdistys. 2014. “The Finnish Evaluation Society.” Accessed 3 January, 2014. <http://www.sayfes.fi/in+english/>.
- Velengtas, Priscilla, Penny Mohr, and Donna A. Messner. 2012. *Making informed decisions: assessing the strengths and weaknesses of study designs and analytic methods for comparative effectiveness research*. Washington: DC: National Pharmaceutical Council.

PART I:
DIFFERENT APPROACHES
TO EVALUATION METHODS AND DESIGNS

CHAPTER ONE

COMPARATIVE EFFECTIVENESS RESEARCH:
DESIGNS AND METHODS
FOR SOCIAL INTERVENTION EVALUATION

EDWARD J MULLEN

Abstract

This chapter examines evaluation of social interventions using comparative effectiveness research designs & methods (CER). While comparative designs and methods have been used for some time to evaluate social interventions, CER is a recent development because of the set of characteristics which are now used to define this form of research. CER has taken on new importance because of the resources allocated to CER, as well as its centrality in policy. CER is presented as an outgrowth of earlier research emphases on technology assessment, evaluation research, outcomes research, and evidence-based policy and practice. After defining CER, this chapter discusses design and methodological requirements and alternatives. Key among these requirements are: the use of comparison research designs contrasting credible, active interventions (not placebo controls); focus on effectiveness rather than efficacy questions; development, expansion, and use of a variety of data sources and methods to conduct timely and relevant research; dissemination of results in a form that is quickly usable by practitioners, clients, policymakers, and payers; conducted in “real-world” settings”; and client involvement in all research phases. With the advent of evidence-based policy and practice, evidence quality came to be viewed through the lens of linear, research design hierarchies. CER has displaced that view with a non-hierarchical view of evidence judged against criteria of relevance and responsiveness to CER requirements. While CER is now prominent in European and American health care, it is only beginning to be discussed in the wider human services. How the CER framework can be adapted to social intervention research is a pressing question.

Introduction

The nature and function of evidence in evidence-based policy and practice is being shaped by many new developments, including the rapid emergence of comparative effectiveness research (CER), which is a response to evidence-demands coming from evidence-based policy and practice. This chapter examines CER and the types and function of evidence required by CER in the context of social intervention evaluation. This is a timely and relevant topic for those concerned with methodological development in the evaluation of social interventions.

This chapter: (1) introduces a research strategy becoming popular in U.S. health care, “comparative effectiveness research” (CER); (2) highlights design requirements for CER; (3) outlines recommended CER research designs and analytic methods; and, (4) identifies issues and limitations needing attention as CER is adapted for social intervention evaluation in the human services.

Emergence of CER

In the United States CER has taken on new importance since passage of the American Recovery and Reinvestment Act of 2009 and the Patient Protection and Affordable Care Act of 2010. (Congressional Budget Office 2007, United States Congress 2009, 2010)

In their introductory editorial to the April 2012 special *JAMA* issue on comparative effectiveness research, Golub and Fontanarosa provide a succinct description of the invigoration of CER in the United States in the context of health care:

Comparative effectiveness research (CER) has captured the attention of the biomedical community, including physicians, other health care professionals, and clinical researchers; the public, including patients and their advocates; and policy makers, including funding agencies and health care insurers. This keen interest is based, at least in part, on the hope that the findings from CER will provide useful information to help clinicians make evidence-based decisions, will incorporate patient preferences and patient-centered perspectives, and, ultimately, will improve the quality of care and help control health care costs. Although the basic concept of comparing available therapies certainly is not new, until recently only a small fraction of US health-related expenditures had been devoted to CER. However, CER achieved prominence when the American Recovery and Reinvestment Act of 2009 allocated \$1.1 billion

to CER: \$300 million in CER funding for the Agency for Healthcare Research and Quality, \$400 million for the National Institutes of Health, and \$400 million for the Office of the Secretary of Health and Human Services. (Golub and Fontanarosa 2012, 1643–1644)

They go on:

Other developments, such as the creation of the Patient-Centered Outcomes Research Institute (PCORI) as part of the Patient Protection and Affordable Care Act of 2010, also have contributed to highlighting the importance of patient involvement in clinical research and in the generation of evidence, thereby bringing CER to the forefront of health policy discussions. (Golub and Fontanarosa 2012, 1643–1644)

While the strongest impetus for CER in the United States has come from the health sector, human service professions in other sectors have turned attention to CER since the passage of the Patient Protection and Affordable Care Act of 2009.

For example, the National Association of Social Workers, Social Work Policy Institute convened a symposium in the autumn of 2009 in Washington, DC for the purpose of strengthening the connection between social work research and CER. (Social Work Policy Institute 2010)

CER is now being included in discussions within public health, social work, and other social intervention disciplines in the context of dissemination and implementation research and translational science. (Glasgow and Steiner 2012)

Since CER is most highly developed in medicine and health this chapter describes CER primarily in the context of health. However, with relevant adaptations, CER has significant potential for applications in the evaluation of social interventions in a wide range of human service areas. I am reminded of how evidence-based policy and practice first developed in medicine and health as evidence-based medicine (EBM), and then quickly found applications in allied health professions and beyond. I believe CER has a similar potential. Indeed some have referred to CER as “EBM-version 2”!

CER Defined

Definitions of CER have been proposed by the U.S. Federal Coordinating Council on Comparative Effectiveness Research, the U.S. Institute of

Medicine (IOM), and the U.S. Agency for Healthcare Research and Quality (AHRQ).

The U.S. Federal Coordinating Council on Comparative Effectiveness Research was established by the U.S. Congress in 2009 charged with defining CER as well as coordinating CER efforts across federal agencies. In the Council's report to the President and to Congress, CER was described as: "the conduct and synthesis of research comparing benefits and harms of different interventions and strategies to prevent, diagnose, treat and monitor health conditions in 'real-world' settings." (Federal Coordinating Council on Comparative Effectiveness Research 2009, 5) It is noteworthy that the Council defined CER as research conducted in real-world settings as contrasted with highly controlled environments wherein efficacy research is conducted. This has important implications which are examined subsequently. Also, this definition refers to the target of interventions and strategies as "health conditions" which is quite general and very inclusive (Federal Coordinating Council for Comparative Effectiveness Research, 2009).

At the time the Federal Coordinating Council was established, the establishing legislation also charged the U.S. Department of Health and Human Services to contract with the U.S. Institute of Medicine both to define CER and to recommend priorities for CER. (IOM (Institute of Medicine) 2009) The Institute defined comparative effectiveness research as:

the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care. The purpose of CER is to assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve health care at both the individual and population levels. (IOM 2009, 41)

This definition expands the target of intervention in an important way, namely, rather than using the general term of "health conditions" as the outcome, this definition explicitly states that the outcomes or targets of intervention can be either a "specific clinical condition" or strategies for "delivery of care". This means that the outcomes can focus on specific individual health problems or aspects of population level service systems. The intent of CER is to improve care at both individual and population levels. Also, the IOM definition importantly adds a statement about the purpose of CER which is not only to "generate and synthesize" (research) evidence, but the definition states that this should be done so as "to assist consumers, clinicians, purchas-